

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2004, 10:34:11 ; Search time 204 Seconds  
(without alignments)  
3373.950 Million cell updates/sec

Title: US-10-088-467-2  
Perfect score: 12668  
Sequence: 1 MGFLHQLQLLLWKNVTLKRR.....GLISFEEERAQLSFNTDTLC 2436

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

		%					
Result		Query					
No.	Score	Match	Length	DB	ID	Description	
1	12668	100.0	2436	4	AAB62210	Aab62210 Human ABC	
2	12668	100.0	2436	5	ABP52093	Abp52093 Homo sapi	
3	12660	99.9	2436	5	ABB76715	Abb76715 Human ATP	
4	12660	99.9	2436	5	AAE22903	Aae22903 Human tra	
5	12656	99.9	2436	4	AAG67160	Aag67160 Amino aci	
6	11725	92.6	2434	5	ABB76716	Abb76716 Rat ATP b	
7	10249	80.9	2001	4	AAAY72649	Aay72649 Human ATP	
8	10249	80.9	2001	5	ABB98347	Abb98347 Human ABC	
9	9237	72.9	1771	5	AAE16781	Aae16781 Human tra	

10	4244.5	33.5	2261	3	AAB38110	Aab38110	Human	ABC
11	4244.5	33.5	2261	4	AAB31363	Aab31363	Amino	aci
12	4244.5	33.5	2261	4	AAB31367	Aab31367	Amino	aci
13	4241.5	33.5	2261	3	AAB38109	Aab38109	Human	ABC
14	4240.5	33.5	2261	3	AAB38117	Aab38117	Human	ABC
15	4240.5	33.5	2261	3	AAB38114	Aab38114	Human	ABC
16	4240.5	33.5	2261	3	AAB38082	Aab38082	Human	ABC
17	4240.5	33.5	2261	3	AAB38112	Aab38112	Human	ABC
18	4240.5	33.5	2261	4	AAB71749	Aab71749	Human	ABC
19	4240.5	33.5	2261	4	AAB31361	Aab31361	Amino	aci
20	4240.5	33.5	2261	4	AAB31365	Aab31365	Amino	aci
21	4240.5	33.5	2261	6	ABU11899	Abu11899	Human	ATP
22	4239.5	33.5	2261	3	AAB38111	Aab38111	Human	ABC
23	4238.5	33.5	2259	3	AAB38107	Aab38107	Human	ABC
24	4238.5	33.5	2261	3	AAB38113	Aab38113	Human	ABC
25	4237.5	33.5	2261	3	AAB38115	Aab38115	Human	ABC
26	4237.5	33.5	2261	6	ABR62034	Abr62034	Human	ABC
27	4237.5	33.5	2261	6	ABR62033	Abr62033	Human	ABC
28	4236.5	33.4	2261	3	AAB38105	Aab38105	Human	ABC
29	4236.5	33.4	2261	4	AAB31362	Aab31362	Amino	aci
30	4236.5	33.4	2261	4	AAB31366	Aab31366	Amino	aci
31	4235.5	33.4	2261	3	AAB38116	Aab38116	Human	ABC
32	4234.5	33.4	2261	5	ABB83115	Abb83115	Polymorph	
33	4233.5	33.4	2256	6	ABR40000	Abr40000	Human	ABC
34	4233.5	33.4	2261	4	AAM50228	Aam50228	Human	ATP
35	4233.5	33.4	2261	4	AAE13022	Aae13022	Human	ATP
36	4233.5	33.4	2261	4	AAU02176	Aau02176	Human	ABC
37	4233.5	33.4	2261	4	AAU02183	Aau02183	Human	ABC
38	4233.5	33.4	2261	4	AAU02181	Aau02181	Human	ABC
39	4233.5	33.4	2261	4	AAU02177	Aau02177	Human	ABC
40	4233.5	33.4	2261	5	AAE23000	Aae23000	Human	ABC
41	4233.5	33.4	2261	5	ABP52092	Abp52092	Homo sapi	
42	4233.5	33.4	2261	5	ABB83122	Abb83122	Polymorph	
43	4233.5	33.4	2261	5	ABB83111	Abb83111	Human	ABC
44	4233.5	33.4	2261	5	ABB83119	Abb83119	Polymorph	
45	4233.5	33.4	2261	7	ADE85895	Ade85895	Human	ATP

# ALIGNMENTS

## RESULT 1

AAB62210

ID AAB62210 standard; protein; 2436 AA.

XX

AC AAB62210;

XX

DT 11-JUN-2001 (first entry)

XX

DE Human ABCA2 transporter protein.

XX

KW ABCA2; transporter protein; gene therapy; cell transport; human.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Domain 22..40



FT /note= "transmebrane domain TM 1"  
 FT Domain 227. .914  
 FT /note= "transmebrane domain TM 6"  
 FT Domain 707. .729  
 FT /note= "transmebrane domain TM 2"  
 FT Domain 750. .772  
 FT /note= "transmebrane domain TM 3"  
 FT Domain 784. .806  
 FT /note= "transmebrane domain TM 4"  
 FT Domain 813. .835  
 FT /note= "transmebrane domain TM 5"  
 FT Region 1007. .1193  
 FT /note= "ATP binding cassette"  
 FT Domain 1457. .1477  
 FT /note= "hydrophic domain HHD"  
 FT Domain 1794. .1815  
 FT /note= "transmebrane domain TM 7"  
 FT Domain 1845. .1867  
 FT /note= "transmebrane domain TM 8"  
 FT Domain 1876. .1898  
 FT /note= "transmebrane domain TM 9"  
 FT Domain 1905. .1927  
 FT /note= "transmebrane domain TM 10"  
 FT Domain 1946. .1968  
 FT /note= "transmebrane domain TM 11"  
 FT Domain 1988. .2010  
 FT /note= "transmebrane domain TM 12"  
 FT Region 2070. .2252  
 FT /note= "ATP binding cassette"  
 XX  
 PN WO200121798-A2.  
 XX  
 PD 29-MAR-2001.  
 XX  
 PF 31-AUG-2000; 2000WO-US040789.  
 XX  
 PR 20-SEP-1999; 99US-0154839P.  
 XX  
 PA (FOXC-) FOX CHASE CANCER CENT.  
 XX  
 PI Tew KD, Vulevic B, Chen Z;  
 XX  
 DR WPI; 2001-257989/26.  
 DR N-PSDB; AAF57452.  
 XX  
 PT New nucleic acid molecule for screening inhibitors of human ABCA2  
 PT mediated transport, encoding a human ABCA2 transporter protein with a  
 PT multi-domain structure including glycosylation and phosphorylation sites.  
 XX  
 PS Claim 6; Fig 7; 68pp; English.  
 XX  
 CC This represents the human ABCA2 transporter protein having a multi-  
 CC domain structure including a number of glycosylation and phosphorylation  
 CC sites, a lipocalin signature motif, nucleotide binding folds having  
 CC walker A and B ATP binding sites, and a number of membrane spanning  
 CC helices. Human ABCA2 transporter polypeptides and nucleic acid encoding  
 CC them are useful for identification, detection and/or molecular

CC characterization of components involved in the transport of molecules  
CC across cell membranes. The nucleic acid is useful as a probe to detect  
CC the presence of and/or expression of genes encoding ABCA2 proteins, and  
CC in gene therapy. A host cell comprising the nucleic acid is useful for  
CC screening compounds that inhibit human ABCA2 mediated transport  
XX

SQ Sequence 2436 AA;

Query Match 100.0%; Score 12668; DB 4; Length 2436;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MGFLHQQLLLWKNVTLKRRSPWVLAFEIFIPLVLFFILLGLRQKKPTISVKEVPFYTAA 60
      |||
Db      1 MGFLHQQLLLWKNVTLKRRSPWVLAFEIFIPLVLFFILLGLRQKKPTISVKEVPFYTAA 60

Qy     61 PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLDRVVEGNLFDPARPSLGSE 120
      |||
Db     61 PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLDRVVEGNLFDPARPSLGSE 120

Qy    121 LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSPNSTAQAL 180
      |||
Db    121 LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSPNSTAQAL 180

Qy    181 LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPALLEQLTC 240
      |||
Db    181 LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPALLEQLTC 240

Qy    241 TPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARFSGLSAELRNQLDVAKVSQQL 300
      |||
Db    241 TPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARFSGLSAELRNQLDVAKVSQQL 300

Qy    301 GLDAPNGSDSSPQAPPPRRLQALLGDLLDAQKVLQDQVDVLSALALLPQGACTGRTPGPP 360
      |||
Db    301 GLDAPNGSDSSPQAPPPRRLQALLGDLLDAQKVLQDQVDVLSALALLPQGACTGRTPGPP 360

Qy    361 ASGAGGAANGTGAGAVMGPNATAEEGAPSAALATPDTLQGQCSAFVQLWAGLQPILCGN 420
      |||
Db    361 ASGAGGAANGTGAGAVMGPNATAEEGAPSAALATPDTLQGQCSAFVQLWAGLQPILCGN 420

Qy    421 NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF 480
      |||
Db    421 NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF 480

Qy    481 AFVGNVTHYAQVWLNISAEIRSFLEQGRLLQHLRWLQQYVAELRLHPEALNLSDELPPA 540
      |||
Db    481 AFVGNVTHYAQVWLNISAEIRSFLEQGRLLQHLRWLQQYVAELRLHPEALNLSDELPPA 540

Qy    541 LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPEESIVNYTLNQAYQD 600
      |||
Db    541 LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPEESIVNYTLNQAYQD 600

Qy    601 NVTVFASVIFQTRKDGLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF 660
      |||
Db    601 NVTVFASVIFQTRKDGLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF 660

Qy    661 VWIQDMMERAIIDTFVGHVVEPGSYVQMFYPYCYTRDDFLFVIEHMMPLCMVISWVYSV 720
```

Db	661	 VWIQDMMERAIIDTFVGHADVVEPGSYVQMFPYPCYTRDDFLFVIEHMMPLCMVISWVYSV	720
Qy	721	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAVFITGFVQLSISVTALTAILKYQVLMH	780
Db	721	 AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAVFITGFVQLSISVTALTAILKYQVLMH	780
Qy	781	SHVVIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAH	840
Db	781	 SHVVIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAH	840
Qy	841	DKITAFEKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTML	900
Db	841	 DKITAFEKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTML	900
Qy	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSV	960
Db	901	 MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSV	960
Qy	961	MEEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTKVYKDDKKLALNKLSNLNLYENQV	1020
Db	961	 MEEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTKVYKDDKKLALNKLSNLNLYENQV	1020
Qy	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRL	1080
Db	1021	 VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRL	1080
Qy	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVG	1140
Db	1081	 TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVG	1140
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Db	1141	 GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Qy	1201	KCCGSPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Db	1201	 KCCGSPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Qy	1261	HVASCLLVSDTSTELSYILPSEAAKKGAFERLFQHLERSLDALHLSSFGIMDTTLEEVL	1320
Db	1261	 HVASCLLVSDTSTELSYILPSEAAKKGAFERLFQHLERSLDALHLSSFGIMDTTLEEVL	1320
Qy	1321	KVSEEDQSLENSEADVKE SRKDVLPGAEGPASGEGHAGNLARCSELTQSQASLQSASSVG	1380
Db	1321	 KVSEEDQSLENSEADVKE SRKDVLPGAEGPASGEGHAGNLARCSELTQSQASLQSASSVG	1380
Qy	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Db	1381	 SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Qy	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQ	1500
Db	1441	 HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQ	1500
Qy	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560

Db	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Qy	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS LPPTA	1620
Db	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS LPPTA	1620
Qy	1621	GPEMWTSAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTDITGHNVS	1680
Db	1621	GPEMWTSAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTDITGHNVS	1680
Qy	1681	EYLLFTSDRFLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Db	1681	EYLLFTSDRFLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Qy	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFII	1800
Db	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFII	1800
Qy	1801	VAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVIILF	1860
Db	1801	VAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVIILF	1860
Qy	1861	VFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGITAT	1920
Db	1861	VFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGITAT	1920
Qy	1921	VATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKM	1980
Db	1921	VATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKM	1980
Qy	1981	KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDQDVASERQR	2040
Db	1981	KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDQDVASERQR	2040
Qy	2041	VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLGVNGAGKTSTFKM	2100
Db	2041	VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLGVNGAGKTSTFKM	2100
Qy	2101	LTGDESTTGGEAFVNGHSLKELLQVQQSLGYCPCDADFELTAREHLQLYTRLRGISW	2160
Db	2101	LTGDESTTGGEAFVNGHSLKELLQVQQSLGYCPCDADFELTAREHLQLYTRLRGISW	2160
Qy	2161	KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK	2220
Db	2161	KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK	2220
Qy	2221	ARRFLWNLILDLIKTRGSVVLTSMSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG	2280
Db	2221	ARRFLWNLILDLIKTRGSVVLTSMSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG	2280
Qy	2281	YMITVRTKSSQSVKDVRFFNRNFPPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSG	2340
Db	2281	YMITVRTKSSQSVKDVRFFNRNFPPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSG	2340
Qy	2341	VLGIEDYSVSQTTLDNVFNFAKKQSDNLEQQETEPSSALQSPLGCLLSLLRPRSAPTEL	2400
Db	2341	VLGIEDYSVSQTTLDNVFNFAKKQSDNLEQQETEPSSALQSPLGCLLSLLRPRSAPTEL	2400

Qy 2401 RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2436  
 |||||  
 Db 2401 RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2436

# RESULT 2

ABP52093

ID ABP52093 standard; protein; 2436 AA.

XX

AC ABP52093;

XX

DT 10-OCT-2002 (first entry)

XX

DE Homo sapiens ABC transporter ABCA2 protein SEQ ID NO:45.

XX

KW ATP-binding cassette transporter; ABC transporter; modulation; D loop;

KW cancer; bacterial infection; fungal infection; protozoal infection;

KW antibacterial; fungicide; protozoacide.

XX

OS Homo sapiens.

XX

PN EP1217066-A1.

XX

PD 26-JUN-2002.

XX

PF 21-DEC-2000; 2000EP-00870316.

XX

PR 21-DEC-2000; 2000EP-00870316.

XX

PA (UYGE-) UNIV GENT.

XX

DR WPI; 2002-550404/59.

XX

PT Modulating activity of ATP-binding cassette (ABC) transporters by

PT influencing dimerization of nucleotide binding domains through use of D

PT loop sequence of an ABC transporter, or its antisense peptide or peptide

PT mimetic.

XX

PS Disclosure; Fig 3; 290pp; English.

XX

CC The present invention describes a method (M1) for modulating the activity

CC of ATP-binding cassette (ABC) transporters by influencing the

CC dimerisation of the nucleotide binding domains comprises using: (a) a

CC polypeptide (polyP) consisting of 5-50 amino acids comprising the D loop

CC sequence of an ABC transporter (ABP52049 to ABP52091); (b) a polyP

CC consisting of the D loop sequence of an ABC transporter; (c) a peptide

CC mimetic or antisense peptide of (a) or (b). ABC transporters have

CC antibacterial, fungicide and protozoacide activities. (M1) is useful for

CC selectively modulating the activity of ABC transporters belonging to the

CC group of multidrug transporter/P-glycoproteins. Bacterial, fungal or

CC protozoal ABC transporters are involved in the infection of a mammal or

CC in the induction of resistance to antibiotics or drugs in a mammal. (M1)

CC is useful for preventing, treating or alleviating diseases associated

CC with functionality of an ABC transporter. ABP52092 to ABP52140 represent

CC ABC transporter proteins given in the exemplification of the present

CC invention

XX

SQ Sequence 2436 AA;

Query Match 100.0%; Score 12668; DB 5; Length 2436;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MGFLHQLQLLLWKNVTLKRRSPWVLAFEIFIPLVLFFILLGLRQKKPTISVKEVPFYTAA 60
          |||
Db      1 MGFLHQLQLLLWKNVTLKRRSPWVLAFEIFIPLVLFFILLGLRQKKPTISVKEVPFYTAA 60

Qy     61 PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLDRVVEEGNLFDPARPSLGSE 120
          |||
Db     61 PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLDRVVEEGNLFDPARPSLGSE 120

Qy    121 LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSPNSTAQAL 180
          |||
Db    121 LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSPNSTAQAL 180

Qy    181 LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPALLEQLTC 240
          |||
Db    181 LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPALLEQLTC 240

Qy    241 TPGSGELGRILTVPEsqKALQGYRDAVCSGQAAARRRFSGLSAELRNQLDVAKVSQQL 300
          |||
Db    241 TPGSGELGRILTVPEsqKALQGYRDAVCSGQAAARRRFSGLSAELRNQLDVAKVSQQL 300

Qy    301 GLDAPNGSDSSPQAPPRRRLQALLGDLLDAQKVLQDQDVLQDQDVLQDQDVLQDQDVL 360
          |||
Db    301 GLDAPNGSDSSPQAPPRRRLQALLGDLLDAQKVLQDQDVLQDQDVLQDQDVLQDQDVL 360

Qy    361 ASGAGGAANGTGAGAVMGPNATAEEGAPSAAALATPDTLQGQCSAFVQLWAGLQPILCGN 420
          |||
Db    361 ASGAGGAANGTGAGAVMGPNATAEEGAPSAAALATPDTLQGQCSAFVQLWAGLQPILCGN 420

Qy    421 NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF 480
          |||
Db    421 NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF 480

Qy    481 AFVGNVTHYAQVWLNISAEIRSFLQGRLOQHRLRWLQYVAELRLHPEALNLSDELPPA 540
          |||
Db    481 AFVGNVTHYAQVWLNISAEIRSFLQGRLOQHRLRWLQYVAELRLHPEALNLSDELPPA 540

Qy    541 LRQDNFSLPSGMALLQQLDIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQD 600
          |||
Db    541 LRQDNFSLPSGMALLQQLDIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQD 600

Qy    601 NVTVFASVIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF 660
          |||
Db    601 NVTVFASVIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF 660

Qy    661 VWIQDMMERAIIDTFVGHQDVVEPGSYVQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSV 720
          |||
Db    661 VWIQDMMERAIIDTFVGHQDVVEPGSYVQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSV 720

Qy    721 AMTIQHIVAEKEHRLKEVMKTMGLNNAVHVVAVFITGFVQLSISVTALTALIKYQVLMH 780
          |||
```

Db	721	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAWFITGQVLSISVTALTAILKYQVLMH	780
Qy	781	SHVVIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAH	840
Db	781	SHVVIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAH	840
Qy	841	DKITAFEKCIASLMSTTAFGLGSKYFALYEAVAGVGIQWHTFSQSPVEGDDFNLLAVTML	900
Db	841	DKITAFEKCIASLMSTTAFGLGSKYFALYEAVAGVGIQWHTFSQSPVEGDDFNLLAVTML	900
Qy	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSV	960
Db	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSV	960
Qy	961	MEEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTQVYKDDKKLALNKLNLNLYENQV	1020
Db	961	MEEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTQVYKDDKKLALNKLNLNLYENQV	1020
Qy	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR	1080
Db	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR	1080
Qy	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVG	1140
Db	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVG	1140
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Db	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Qy	1201	KCCGSPLFLKGTGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Db	1201	KCCGSPLFLKGTGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Qy	1261	HVASCLLVSDTSTELSYILPSEAAKKGAERLRFQHLERSLDALHLSSFGMDTTLEEVFL	1320
Db	1261	HVASCLLVSDTSTELSYILPSEAAKKGAERLRFQHLERSLDALHLSSFGMDTTLEEVFL	1320
Qy	1321	KVSEEDQSLENSEADVKEERKDVLPAGEPASGEGHAGNLARCSELTQSQASLQSASSVG	1380
Db	1321	KVSEEDQSLENSEADVKEERKDVLPAGEPASGEGHAGNLARCSELTQSQASLQSASSVG	1380
Qy	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Db	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Qy	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQ	1500
Db	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQ	1500
Qy	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Db	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Qy	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS LPPTA	1620
Db	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS LPPTA	1620

Qy	1621	GPEMWTSA <del>PSLPRLVREPVRCTCSA</del> QGTGFSC <del>PSSVGGHPPQMRVVTGDILTDITGHNVS</del>	1680
Db	1621	GPEMWTSA <del>PSLPRLVREPVRCTCSA</del> QGTGFSC <del>PSSVGGHPPQMRVVTGDILTDITGHNVS</del>	1680
Qy	1681	EYLLFTSD <del>RFLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM</del>	1740
Db	1681	EYLLFTSD <del>RFLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM</del>	1740
Qy	1741	PTYLNSLN <del>NAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFII</del>	1800
Db	1741	PTYLNSLN <del>NAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFII</del>	1800
Qy	1801	VAMSFVPAS <del>FVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMLNYLVPATCCVILF</del>	1860
Db	1801	VAMSFVPAS <del>FVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMLNYLVPATCCVILF</del>	1860
Qy	1861	VFDLPAYTS <del>PTNFPAVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGITAT</del>	1920
Db	1861	VFDLPAYTS <del>PTNFPAVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGITAT</del>	1920
Qy	1921	VATFLLQL <del>FEHDKDLKVVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKM</del>	1980
Db	1921	VATFLLQL <del>FEHDKDLKVVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKM</del>	1980
Qy	1981	KSPFEWDIV <del>TRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDDVDASERQR</del>	2040
Db	1981	KSPFEWDIV <del>TRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDDVDASERQR</del>	2040
Qy	2041	VLRGDADND <del>VMVKIENLTKVYKSRKIGRILAVDRCLGVRPGECFGLLGVNAGAKTSTFKM</del>	2100
Db	2041	VLRGDADND <del>VMVKIENLTKVYKSRKIGRILAVDRCLGVRPGECFGLLGVNAGAKTSTFKM</del>	2100
Qy	2101	LTGDESTTG <del>GGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW</del>	2160
Db	2101	LTGDESTTG <del>GGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW</del>	2160
Qy	2161	KDEARVVK <del>WALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK</del>	2220
Db	2161	KDEARVVK <del>WALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK</del>	2220
Qy	2221	ARRFLWNL <del>IILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG</del>	2280
Db	2221	ARRFLWNL <del>IILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG</del>	2280
Qy	2281	YMITVRTK <del>SSQSVKDVVRFFNRNFP EAM LKERHHTKVQYQLKSEHISLAQVFSKMEQVSG</del>	2340
Db	2281	YMITVRTK <del>SSQSVKDVVRFFNRNFP EAM LKERHHTKVQYQLKSEHISLAQVFSKMEQVSG</del>	2340
Qy	2341	VLGIEDYS <del>VSQTTLDNVFVNFAKKQSDNLEQQETEP PSALQSPLGCLLSLLRPRSAPTEL</del>	2400
Db	2341	VLGIEDYS <del>VSQTTLDNVFVNFAKKQSDNLEQQETEP PSALQSPLGCLLSLLRPRSAPTEL</del>	2400
Qy	2401	RALVADEP <del>EDLDTEDEGLISFEEERAQLSFNTDTLC</del>	2436
Db	2401	RALVADEP <del>EDLDTEDEGLISFEEERAQLSFNTDTLC</del>	2436



RESULT 3

ABB76715

ID ABB76715 standard; protein; 2436 AA.  
 XX  
 AC ABB76715;  
 XX  
 DT 06-JUN-2002 (first entry)  
 XX  
 DE Human ATP binding cassette transporter protein, ABCA2.  
 XX  
 KW Human; ABCA2; neuroprotective; nootropic; antiparkinsonian;  
 KW adenosine triphosphate binding cassette transporter protein;  
 KW ATP binding cassette transporter protein; Alzheimer's disease;  
 KW prion disease; Huntington's disease; Parkinson's disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200208424-A1.  
 XX  
 PD 31-JAN-2002.  
 XX  
 PF 26-JUL-2001; 2001WO-JP006457.  
 XX  
 PR 26-JUL-2000; 2000JP-00225462.  
 XX  
 PA (BANY ) BANYU PHARM CO LTD.  
 PA (INAG/) INAGAKI N.  
 XX  
 PI Inagaki N;  
 XX  
 DR WPI; 2002-179907/23.  
 DR N-PSDB; ABL53009.  
 XX  
 PT Adenosine triphosphate (ATP) binding cassette transporter gene ABCA2 of  
 PT human or rat origin and encoded protein, useful for screening inhibitors,  
 PT promoters and regulators of ABCA2 activity as drugs and diagnosis of  
 PT ABCA2-related diseases.  
 XX  
 PS Claim 1; Page 52-64; 118pp; Japanese.  
 XX  
 CC The present sequence is the protein sequence for human adenosine  
 CC triphosphate (ATP) binding cassette transporter protein (ABCA2). ABCA2  
 CC can be used in the diagnosis, treatment and prevention of diseases such  
 CC as Alzheimer's disease, prion diseases, Huntington's disease, and  
 CC Parkinson's disease  
 XX  
 SQ Sequence 2436 AA;

Query Match 99.9%; Score 12660; DB 5; Length 2436;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2435; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MGFLHQLQLLLWKNVTLKRRSPWVLA FEIFIPLVLF FILLGLRQKKPTISVKEVPFYTAA 60  
 |||  
 Db 1 MGFLHQLQLLLWKNVTLKRRSPWVLA FEIFIPLVLF FILLGLRQKKPTISVKEVSFYTAA 60

Qy	61	PLTSAGILPVMQSLCPDQORDEFGFLQYANSTVTQLLERLDRVVEEGNLFDPARPSLGSE	120
Db	61	PLTSAGILPVMQSLCPDQORDEFGFLQYANSTVTQLLERLDRVVEEGNLFDPARPSLGSE	120
Qy	121	LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSLPNSTAQAL	180
Db	121	LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSLPNSTAQAL	180
Qy	181	LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPALLEQLTC	240
Db	181	LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPALLEQLTC	240
Qy	241	TPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAELRNQLDVAKVSQQ	300
Db	241	TPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAELRNQLDVAKVSQQ	300
Qy	301	GLDAPNGSDSSPQAPPPRRLQALLGDLLDAQKVLQDQVDVLSALALLPQGACTGRTPGPP	360
Db	301	GLDAPNGSDSSPQAPPPRRLQALLGDLLDAQKVLQDQVDVLSALALLPQGACTGRTPGPP	360
Qy	361	ASGAGGAANGTGAGAVMGNATAEEGAPSAAALATPDTLQGQCSAFVQLWAGLQPILCGN	420
Db	361	ASGAGGAANGTGAGAVMGNATAEEGAPSAAALATPDTLQGQCSAFVQLWAGLQPILCGN	420
Qy	421	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF	480
Db	421	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF	480
Qy	481	AFVGNVTHYAQVWLNISAEIRSFLQGRLLQQLRWLQQYVAELRLHPEALNLSDELPPA	540
Db	481	AFVGNVTHYAQVWLNISAEIRSFLQGRLLQQLRWLQQYVAELRLHPEALNLSDELPPA	540
Qy	541	LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQD	600
Db	541	LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQD	600
Qy	601	NVTVFASVIFQTRKDGSLLPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF	660
Db	601	NVTVFASVIFQTRKDGSLLPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF	660
Qy	661	VWIQDMMERAIIDTFVGHDVVEPGSYVQMFYPYCYTRDDFLFVIEHMMPLCMVISWVYSV	720
Db	661	VWIQDMMERAIIDTFVGHDVVEPGSYVQMFYPYCYTRDDFLFVIEHMMPLCMVISWVYSV	720
Qy	721	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAWFITGFVQLSISVTALTALIKYQVLMH	780
Db	721	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAWFITGFVQLSISVTALTALIKYQVLMH	780
Qy	781	SHVVIWFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAH	840
Db	781	SHVVIWFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAH	840
Qy	841	DKITAFEKCIASLMSTTAFGLGSKYFALYEAVAGVGIQWHTFSQSPVEGDDFNLLAVTML	900
Db	841	DKITAFEKCIASLMSTTAFGLGSKYFALYEAVAGVGIQWHTFSQSPVEGDDFNLLAVTML	900
Qy	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSV	960

Db	901	MVDVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSV	960
Qy	961	MEEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTKVYKDDKKLALNKLNLNLYENQV	1020
Db	961	MEEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTKVYKDDKKLALNKLNLNLYENQV	1020
Qy	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRL	1080
Db	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRL	1080
Qy	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVG	1140
Db	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVG	1140
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Db	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Qy	1201	KCCGSPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Db	1201	KCCGSPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Qy	1261	HVASCLLVSDTSTELSYILPSEAAKGAFERLFQHLERSLDALHLSSFGLMDTTLEEVFL	1320
Db	1261	HVASCLLVSDTSTELSYILPSEAAKGAFERLFQHLERSLDALHLSSFGLMDTTLEEVFL	1320
Qy	1321	KVSEEDQSLENSEADVKEERKDVLPAGEPASGEGHAGNLARCELTQSQASLQSASSVG	1380
Db	1321	KVSEEDQSLENSEADVKEERKDVLPAGEPASGEGHAGNLARCELTQSQASLQSASSVG	1380
Qy	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALS RVGQGSRKLDGGWLKVRQF	1440
Db	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALS RVGQGSRKLDGGWLKVRQF	1440
Qy	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQ	1500
Db	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQ	1500
Qy	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Db	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Qy	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDELQAWNVS LPPTA	1620
Db	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDELQAWNVS LPPTA	1620
Qy	1621	GPEMWTSAPSLPRLVREPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILTDTIGHNVS	1680
Db	1621	GPEMWTSAPSLPRLVREPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILTDTIGHNVS	1680
Qy	1681	EYLLFTSDRFRHLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Db	1681	EYLLFTSDRFRHLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Qy	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFII	1800

Db 1741 PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFII 1800

QY 1801 VAMSFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMLNLYLPATCCVIILF 1860  
 |||

Db 1801 VAMSFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMLNLYLPATCCVIILF 1860

QY 1861 VFDDLPAYTSPTNFPAVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITAT 1920  
 |||

Db 1861 VFDDLPAYTSPTNFPAVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITAT 1920

QY 1921 VATFLLQLFEHDKDLKVVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDM 1980  
 |||

Db 1921 VATFLLQLFEHDKDLKVVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDM 1980

QY 1981 KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMFVSTKPVEDDQDVASERQR 2040  
 |||

Db 1981 KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMFVSTKPVEDDQDVASERQR 2040

QY 2041 VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLGVNGAGKTSTFKM 2100  
 |||

Db 2041 VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLGVNGAGKTSTFKM 2100

QY 2101 LTGDESTTGGEAFVNGHSLVKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW 2160  
 |||

Db 2101 LTGDESTTGGEAFVNGHSLVKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW 2160

QY 2161 KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK 2220  
 |||

Db 2161 KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK 2220

QY 2221 ARRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG 2280  
 |||

Db 2221 ARRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG 2280

QY 2281 YMITVRTKSSQSVKDVVRFFNRFPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSG 2340  
 |||

Db 2281 YMITVRTKSSQSVKDVVRFFNRFPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSG 2340

QY 2341 VLGIEDYSVSQTTLDNVFVNFAKKQSDNLEQQETEPSSALQSPLGCLLSLLRPRSAPTEL 2400  
 |||

Db 2341 VLGIEDYSVSQTTLDNVFVNFAKKQSDNLEQQETEPSSALQSPLGCLLSLLRPRSAPTEL 2400

QY 2401 RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2436  
 |||

Db 2401 RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2436

RESULT 4

AAE22903

ID AAE22903 standard; protein; 2436 AA.

XX

AC AAE22903;

XX

DT 09-AUG-2002 (first entry)

XX

DE Human transporter and ion channel (TRICH) 2.

XX

KW Human; transporter and ion channel; TRICH; transport disorder;  
KW diabetes mellitus; angina; Alzheimer's disease; neurological; epilepsy;  
KW stroke; Huntington's disease; meningitis; muscle; myocarditis; cancer;  
KW infectious myositis; arrhythmia; asthma; immunological; gene therapy;  
KW acquired immunodeficiency syndrome; AIDS; allergy; atherosclerosis;  
KW cell proliferative disorder; cerebroprotective; cirrhosis; hepatitis;  
KW transgenic; neuroprotective; anticonvulsant; nootropic; cytostatic;  
KW antiinflammatory; hepatotropic; psoriasis.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
FT	Domain	22. .45
FT		/label= Transmembrane_domain
FT	Domain	784. .803
FT		/label= Transmembrane_domain
FT	Domain	893. .911
FT		/label= Transmembrane_domain
FT	Domain	1018. .1198
FT		/note= "ABC transporter domain"
FT	Binding-site	1025. .1032
FT		/note= "ATP/GTP binding site"
FT	Domain	1124. .1138
FT		/note= "ABC transporter motif"
FT	Domain	1424. .1437
FT		/note= "Lipocalin motif"
FT	Domain	1426. .1437
FT		/note= "Lipocalin motif"
FT	Domain	1793. .1813
FT		/label= Transmembrane_domain
FT	Domain	1845. .1862
FT		/label= Transmembrane_domain
FT	Domain	1900. .1926
FT		/label= Transmembrane_domain
FT	Domain	2081. .2262
FT		/note= "ABC transporter domain"
FT	Binding-site	2088. .2095
FT		/note= "ATP/GTP binding site"

XX

PN WO200222684-A2.

XX

PD 21-MAR-2002.

XX

PF 14-SEP-2001; 2001WO-US028938.

XX

PR 15-SEP-2000; 2000US-0232685P.

PR 22-SEP-2000; 2000US-0234842P.

PR 29-SEP-2000; 2000US-0236882P.

PR 05-OCT-2000; 2000US-0239057P.

PR 13-OCT-2000; 2000US-0240540P.

PR 18-OCT-2000; 2000US-0241700P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Lee EA, Yue H, Lal PG, Walia NK, Baughn MR, Warren BA, Lee S;  
PI Sanjanwala MS, Yao MG, Ramkumar J, Thornton M, Gandhi AR;  
PI Policky JL, Elliott VS, Arvizu C, Raumann BE, Bruns CM, Naini A;



Db	241	TPGSSELGRILTVPESEQKALQGYRDAVCSGQAAARARRFSGLSAELRNQLDVAKVSQQL	300
Qy	301	GLDAPNGSDSSPQAPPPRRLLQALLGDLLDAQVQLQDQVLSALALLLPQGACTGRTPGPP	360
Db	301	GLDAPNGSDSSPQAPPPRRLLQALLGDLLDAQVQLQDQVLSALALLLPQGACTGRTPGPP	360
Qy	361	ASGAGGAANGTGAGAVMGP NATAE EGAPSAAALATPDTLQGQCSAFVQLWAGLQPILCGN	420
Db	361	ASGAGGAANGTGAGAVMGP NATAE EGAPSAAALATPDTLQGQCSAFVQLWAGLQPILCGN	420
Qy	421	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF	480
Db	421	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF	480
Qy	481	AFVGNVTHYAQVWLNISAEIRSFLQGRLLQHLRWLQQYVAELRLHPEALNLSLDELPPA	540
Db	481	AFVGNVTHYAQVWLNISAEIRSFLQGRLLQHLRWLQQYVAELRLHPEALNLSLDELPPA	540
Qy	541	LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPDEESIVNYTLNQAYQD	600
Db	541	LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPDEESIVNYTLNQAYQD	600
Qy	601	NVTVFASVIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF	660
Db	601	NVTVFASVIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF	660
Qy	661	VWIQDMMERAIIDTFVGHDVVEPGSYVQMFYPYCYTRDDFLFVIEHMMPLCMVISWVYSV	720
Db	661	VWIQDMMERAIIDTFVGHDVVEPGSYVQMFYPYCYTRDDFLFVIEHMMPLCMVISWVYSV	720
Qy	721	AMTIQHIVA EKEHRLKEVMKTMGLNNAVHWVAVFITGFVQLSISVTALTAILKYGQVLMH	780
Db	721	AMTIQHIVA EKEHRLKEVMKTMGLNNAVHWVAVFITGFVQLSISVTALTAILKYGQVLMH	780
Qy	781	SHVVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEV AH	840
Db	781	SHVVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEV AH	840
Qy	841	DKITAF EKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTML	900
Db	841	DKITAF EKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTML	900
Qy	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSV	960
Db	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSV	960
Qy	961	MEEDQACAMESRRFEETRGMEEEP THLPLVVCVDKLTKVYKDDKKLALNKL SLNLYENQV	1020
Db	961	MEEDQACAMESRRFEETRGMEEEP THLPLVVCVDKLTKVYKDDKKLALNKL SLNLYENQV	1020
Qy	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR L	1080
Db	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR L	1080
Qy	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAFVG	1140

Db	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVG	1140
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Db	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Qy	1201	KCCGSPLFLKGTYG DG YRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Db	1201	KCCGSPLFLKGTYG DG YRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Qy	1261	HVASCLLVSDTSTELSYILPSEAAKKGA FERLFQH LERSLDALHLSSFGLMDTTLEEVFL	1320
Db	1261	HVASCLLVSDTSTELSYILPSEAAKKGA FERLFQH LERSLDALHLSSFGLMDTTLEEVFL	1320
Qy	1321	KVSEEDQSLSENSEADV KESRKDVLPGAEGPASGEGHAGNLARCSELTSQSASLQSASSVG	1380
Db	1321	KVSEEDQSLSENSEADV KESRKDVLPGAEGPASGEGHAGNLARCSELTSQSASLQSASSVG	1380
Qy	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Db	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Qy	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQ	1500
Db	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQ	1500
Qy	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Db	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Qy	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS LPPTA	1620
Db	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS LPPTA	1620
Qy	1621	GPEMWTSAPSLPRLVREPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILT DITGH NVS	1680
Db	1621	GPEMWTSAPSLPRLVREPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILT DITGH NVS	1680
Qy	1681	EYLLFTSDRFLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Db	1681	EYLLFTSDRFLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Qy	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHMPMNKTSASLSLDYLLQGT DVVIAIFII	1800
Db	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHMPMNKTSASLSLDYLLQGT DVVIAIFII	1800
Qy	1801	VAMSFVPASFVVFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVIILF	1860
Db	1801	VAMSFVPASFVVFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVIILF	1860
Qy	1861	VFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWFEPSSAYVFLIVINLFIGITAT	1920
Db	1861	VFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWFEPSSAYVFLIVINLFIGITAT	1920
Qy	1921	VATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKM	1980
Db	1921	VATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKM	1980



Qy 1981 KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDQDVASERQR 2040  
 |||||  
 Db 1981 KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDQDVASERQR 2040

Qy 2041 VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRCLGVPRGECFGLLVNGAGKTSTFKM 2100  
 |||||  
 Db 2041 VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRCLGVPRGECFGLLVNGAGKTSTFKM 2100

Qy 2101 LTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW 2160  
 |||||  
 Db 2101 LTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW 2160

Qy 2161 KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK 2220  
 |||||  
 Db 2161 KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK 2220

Qy 2221 ARRFLWNLILDLIKTRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG 2280  
 |||||  
 Db 2221 ARRFLWNLILDLIKTRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG 2280

Qy 2281 YMITVVRTKSSQSVKDVVRFFNRNFPPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSG 2340  
 |||||  
 Db 2281 YMITVVRTKSSQSVKDVVRFFNRNFPPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSG 2340

Qy 2341 VLGIEDYSVSQTTLDNVFNFAKKQSDNLEQQETEPFSPALQSPLGCLLSLLRPRSAPTEL 2400  
 |||||  
 Db 2341 VLGIEDYSVSQTTLDNVFNFAKKQSDNLEQQETEPFSPALQSPLGCLLSLLRPRSAPTEL 2400

Qy 2401 RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2436  
 |||||  
 Db 2401 RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2436

# RESULT 5

AAG67160

ID AAG67160 standard; protein; 2436 AA.

XX

AC AAG67160;

XX

DT 13-NOV-2001 (first entry)

XX

DE Amino acid sequence of a human 17114 transporter polypeptide.

XX

KW Human; transporter; 20685; 579; 17114; 23821; 33894; 32613;

KW vesicular monoamine transporter; neurotransmitter-symporter;

KW ABC transporter; sulfate transporter; neurological disorder;

KW central nervous system disorder; Parkinson's disease; depression; pain;

KW infectious disease; cell proliferative disorder; cancer; blood disorder;

KW immune disorder; inflammatory disorder; spleen disorder; lung disorder;

KW Hodgkin's disease; Niemann-Pick disease; chronic bronchitis; ischemia;

KW colon disorder; cirrhosis; uterus disorder; endometrium disorder;

KW endometrial stromal tumour; brain disorder; T-cell disorder; anemia;

KW Sjogren syndrome; skin disorder; lupus erythematosus; heart disorder;

KW haematopoietic stem cell; Alzheimer's disease; myocardial infarction;

KW blood vessel; Kawasaki syndrome; red cell disorder; thymus disorder;

KW B-cell disorder; kidney disorder; glomerulonephritis; breast disorder;

KW testis disorder; thyroid disorder; Graves disease; pancreatitis;  
 KW skeletal muscle disorder; tumour; pancreas disorder;  
 KW small intestine disorder; celiac sprue.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 23. .42  
 FT /note= "transmembrane domain"  
 FT Domain 54. .71  
 FT /note= "transmembrane domain"  
 FT Domain 707. .724  
 FT /note= "transmembrane domain"  
 FT Domain 750. .772  
 FT /note= "transmembrane domain"  
 FT Domain 783. .806  
 FT /note= "transmembrane domain"  
 FT Domain 813. .834  
 FT /note= "transmembrane domain"  
 FT Domain 893. .914  
 FT /note= "transmembrane domain"  
 FT Domain 1018. .1198  
 FT /note= "ABC transporter domain"  
 FT Domain 1457. .1479  
 FT /note= "transmembrane domain"  
 FT Domain 1793. .1816  
 FT /note= "transmembrane domain"  
 FT Domain 1846. .1862  
 FT /note= "transmembrane domain"  
 FT Domain 1875. .1898  
 FT /note= "transmembrane domain"  
 FT Domain 1905. .1929  
 FT /note= "transmembrane domain"  
 FT Domain 2081. .2262  
 FT /note= "ABC transporter domain"  
 XX  
 PN WO200164875-A2.  
 XX  
 PD 07-SEP-2001.  
 XX  
 PF 28-FEB-2001; 2001WO-US006374.  
 XX  
 PR 29-FEB-2000; 2000US-0185906P.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Glucksmann MA;  
 XX  
 DR WPI; 2001-550178/61.  
 DR N-PSDB; AAH75187.  
 XX  
 PT Novel human transporter polypeptides useful for treating and diagnosing  
 PT Parkinson's disease, Hodgkin disease, glomerulonephritis, myocardial  
 PT infarction, Grave's disease, Alzheimer's disease, anemia, asthma and  
 PT tumors.  
 XX  
 PS Claim 9; Fig 14A-G; 259pp; English.

XX

CC The present sequence represents a human transporter polypeptide. The  
CC specification describes 20685, 579, 17114, 23821, 33894 or 32613 human  
CC transporter polypeptides. The 20685 transporter is similar to vesicular  
CC monoamine transporters. The 579 transporter is similar to  
CC neurotransmitter-symporters. The 17114 transporter is similar to ABC  
CC transporters. The 32613 transporter is similar to sulfate transporters.  
CC The transporter polypeptides and polynucleotides are useful for treating  
CC and diagnosing neurological and central nervous system disorders (e.g.  
CC Parkinson's disease, depression, pain), infectious disease, cell  
CC proliferative disorders (e.g., cancer), blood disorders, and immune and  
CC inflammatory disorders. They are also useful for treating and diagnosing  
CC disorders involving the spleen (e.g., Hodgkin disease, Niemann-Pick  
CC disease), lung (e.g., chronic bronchitis), colon (cirrhosis), uterus and  
CC endometrium (e.g., endometrial stromal tumours), brain (e.g., ischemia),  
CC T-cells (e.g., Sjogren syndrome), skin (lupus erythematosus),  
CC haematopoietic stem cells (e.g, Alzheimer's disease), heart (e.g.,  
CC myocardial infarction), blood vessels (e.g., Kawasaki syndrome), red  
CC cells (e.g., anemias), disorders involving thymus, B-cells, kidney (e.g.,  
CC glomerulonephritis), disorders involving breast, testis, epididymis,  
CC prostate, thyroid (e.g., Graves disease), disorders involving skeletal  
CC muscle (e.g, tumour), pancreas (e.g., pancreatitis), small intestine  
CC (e.g., celiac sprue), disorders related to reduced platelet number and  
CC ovary

XX

SQ Sequence 2436 AA;

Query Match 99.9%; Score 12656; DB 4; Length 2436;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2434; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 MGFLHQLQLLLWKNVTCLKRRSPWVLAFEIFIPLVLFILLGLRQKKPTISVKEVPFYTAA 60
          |||
Db      1 MGFLHQLQLLLWKNVTCLKRRSPWVLAFEIFIPLVLFILLGLRQKKPTISVKEVSFYTAA 60

Qy     61 PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLDRVVEEGNLFDPARPSLGSE 120
          |||
Db     61 PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLDRVVEEGNLFDPARPSLGSE 120

Qy    121 LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDVARNPQELWRFLTQNLSLPNSTAQAL 180
          |||
Db    121 LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDVARNPQELWRFLTQNLSLPNSTAQAL 180

Qy    181 LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPALLEQLTC 240
          |||
Db    181 LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPALLEQLTC 240

Qy    241 TPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARFSGLSAELRNQLDVAKVSQQL 300
          |||
Db    241 TPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARFSGLSAELRNQLDVAKVSQQL 300

Qy    301 GLDAPNGSDSSPQAPPPRRLQALLGDLLDAQKVLQDQVDVLSALALLLPQGACTGRTPGPP 360
          |||
Db    301 GLDAPNGSDSSPQAPPPRRLQALLGDLLDAQKVLQDQVDVLSALALLLPQGACTGRTPGPP 360

Qy    361 ASGAGGAANGTGAGAVMGPNATAEEGAPSAAALATPDTLQGQCSAFVQLWAGLQPILCGN 420
          |||
```

Db	361	ASGAGGAANGTGAGAVMGPNATAEEGAPSAAALATPDTLQGQCSAFVQLWAGLQPILCGN	420
Qy	421	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF	480
Db	421	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF	480
Qy	481	AFVGNVTHYAQVWLNISAEIRSFLEQGRLQQHLRWLQQYVAELRLHPEALNLSDELPPA	540
Db	481	AFVGNVTHYAQVWLNISAEIRSFLEQGRLQQHLRWLQQYVAELRLHPEALNLSDELPPA	540
Qy	541	LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPEESIVNYTLNQAYQD	600
Db	541	LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPEESIVNYTLNQAYQD	600
Qy	601	NVTVFASVIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF	660
Db	601	NVTVFASVIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF	660
Qy	661	VWIQDMMERAIIDTFVGHDVVEPGSYVQMFYPYCYTRDDFLFVIEHMMPLCMVISWVYSV	720
Db	661	VWIQDMMERAIIDTFVGHDVVEPGSYVQMFYPYCYTRDDFLFVIEHMMPLCMVISWVYSV	720
Qy	721	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAWFITGFVQLSISVTALTAILKYGQVLMH	780
Db	721	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAWFITGFVQLSISVTALTAILKYGQVLIH	780
Qy	781	SHVVIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIFLSYVPYMYVAIREEVAH	840
Db	781	SHVVIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIFLSYVPYMYVAIREEVAH	840
Qy	841	DKITAFEKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLLAVTML	900
Db	841	DKITAFEKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLLAVTML	900
Qy	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSV	960
Db	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSV	960
Qy	961	MEEDQACAMESRRFEETRGMEEEPTHLPLVVCVDKLTKVYKDDKKLALNKLNLNLYENQV	1020
Db	961	MEEDQACAMESRRFEETRGMEEEPTHLPLVVCVDKLTKVYKDDKKLALNKLNLNLYENQV	1020
Qy	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR	1080
Db	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR	1080
Qy	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAFVG	1140
Db	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAFVG	1140
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Db	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Qy	1201	KCCGSPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Db	1201	KCCGSPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260

Qy	1261	HVASCLLVSDTSTELSYILPSEAAKKGA FERLFQHLERSLDALHLSSFGLMDTTLEEVFL	1320
Db	1261	HVASCLLVSDTSTELSYILPSEAAKKGA FERLFQHLERSLDALHLSSFGLMDTTLEEVFL	1320
Qy	1321	KVSEEDQSLSENSEADV KESRKDVLP GAEGPASGEGHAGNIARCSELTQSQASLQSASSVG	1380
Db	1321	KVSEEDQSLSENSEADV KESRKDVLP GAEGPASGEGHAGNIARCSELTQSQASLQSASSVG	1380
Qy	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Db	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Qy	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVP EIGDLPPLVLSPSQYHNYTQ	1500
Db	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVP EIGDLPPLVLSPSQYHNYTQ	1500
Qy	1501	PRGNFI PYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Db	1501	PRGNFI PYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Qy	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS LPPTA	1620
Db	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS LPPTA	1620
Qy	1621	GPEMWT SAPSLPRLVREPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILT DITGHNV	1680
Db	1621	GPEMWT SAPSLPRLVREPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILT DITGHNV	1680
Qy	1681	EYLLFTSDRFLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Db	1681	EYLLFTSDRFLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Qy	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGT DVVIAIFII	1800
Db	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGT DVVIAIFII	1800
Qy	1801	VAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVIILF	1860
Db	1801	VAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVIILF	1860
Qy	1861	VFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITAT	1920
Db	1861	VFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITAT	1920
Qy	1921	VATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKM	1980
Db	1921	VATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKM	1980
Qy	1981	KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRM PVSTKPVEDDVDVASERQR	2040
Db	1981	KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRM PVSTKPVEDDVDVASERQR	2040
Qy	2041	VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFLLGVNGAGKTSTFKM	2100
Db	2041	VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFLLGVNGAGKTSTFKM	2100

Qy	2101	LTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW	2160
Db	2101	LTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW	2160
Qy	2161	KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK	2220
Db	2161	KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK	2220
Qy	2221	ARRFLWNLILDLIKTRGSVVLTSMSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG	2280
Db	2221	ARRFLWNLILDLIKTRGSVVLTSMSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG	2280
Qy	2281	YMITVRTKSSQSVKDVVRFNRFPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSG	2340
Db	2281	YMITVRTKSSQSVKDVVRFNRFPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSG	2340
Qy	2341	VLGIEDYSVSQTTLDNVFVNFPAKQSDNLEQQETEPSPALQSPLGCLLSLLRPRSAPTL	2400
Db	2341	VLGIEDYSVSQTTLDNVFVNFPAKQSDNLEQQETEPSPALQSPLGCLLSLLRPRSAPTL	2400
Qy	2401	RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC	2436
Db	2401	RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC	2436

RESULT 6

ABB76716

ID ABB76716 standard; protein; 2434 AA.

XX

AC ABB76716;

XX

DT 06-JUN-2002 (first entry)

XX

DE Rat ATP binding cassette transporter protein, ABCA2.

XX

KW Rat; ABCA2; neuroprotective; nootropic; antiparkinsonian;  
 KW adenosine triphosphate binding cassette transporter protein;  
 KW ATP binding cassette transporter protein; Alzheimer's disease;  
 KW prion disease; Huntington's disease; Parkinson's disease.

XX

OS Rattus sp.

XX

PN WO200208424-A1.

XX

PD 31-JAN-2002.

XX

PF 26-JUL-2001; 2001WO-JP006457.

XX

PR 26-JUL-2000; 2000JP-00225462.

XX

PA (BANY ) BANYU PHARM CO LTD.

PA (INAG/) INAGAKI N.

XX

PI Inagaki N;

XX

DR WPI; 2002-179907/23.

DR N-PSDB; ABL53011.

XX  
PT Adenosine triphosphate (ATP) binding cassette transporter gene ABCA2 of  
PT human or rat origin and encoded protein, useful for screening inhibitors,  
PT promoters and regulators of ABCA2 activity as drugs and diagnosis of  
PT ABCA2-related diseases.  
XX  
PS Claim 6; Page 87-99; 118pp; Japanese.  
XX  
CC The present sequence is the protein sequence for rat adenosine  
CC triphosphate (ATP) binding cassette transporter protein (ABCA2). ABCA2  
CC can be used in the diagnosis, treatment and prevention of diseases such  
CC as Alzheimer's disease, prion diseases, Huntington's disease, and  
CC Parkinson's disease  
XX  
SQ Sequence 2434 AA;

Query Match 92.6%; Score 11725; DB 5; Length 2434;  
Best Local Similarity 92.8%; Pred. No. 0;  
Matches 2262; Conservative 49; Mismatches 122; Indels 4; Gaps 4;

Qy	1	MGFLHQQLLLWKNVTLKRRSPWVLAFEIFIPLVLFILLGLRQKKPTISVKEVPFYTAA	60
Db	1	MGFLHQQLLLWKNVTLKRRSPWVLAFEIFIPLVLFILLGLRQKKPTISVKEA-FYTAA	59
Qy	61	PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLDRVVEEGNLFDPARPSLGSE	120
		:	
Db	60	PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLNRVVEESNLFDPERPSLGSE	119
Qy	121	LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSLPNSTAQAL	180
		:	
Db	120	LEALHQRLEALSSGPGTWESHARPVSSFSLSVARDKRELWRFLMQNLSPNSTAQAL	179
Qy	181	LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPALLEQLTC	240
		:	
Db	180	LAARVDPSEVYRLLFGPLPDLDGKLGFLRKQEPWSHLSNPLFQMEELLAPALLEQLTC	239
Qy	241	TPGSSELGRILTVPEQKQALQGYRDAVCSGQAAARARRFSGLSAELRNQLDVAKVSQQL	300
		:	
Db	240	APGSSELGRILTMPEGHQVDLQGYRDAVCSGQATARAQHFSDLATELRNQLDIAKIAQQL	299
Qy	301	GLDAPNGSDSSPQAPPPRRLQALLGDLLDAQKVLQDQVDVLSALALLLPQGACTGRTPGPP	360
		:     :	
Db	300	GFNVPNGSDPQPAPSPQSLQALLGDLLDVQKVLQDQVDVLSALALLLPQGACAGRAPAPQ	359
Qy	361	ASGAGGAANGTGAGAVMGPNATAEEGAPSAALATPDTLQGQCSAFVQLWAGLQPILCGN	420
		:	
Db	360	AGSPSGPANSTGVGANTGPNTTVEEGTQSPVTPASPDTLQGQCSAFVQLWAGLQPILCGN	419
Qy	421	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF	480
Db	420	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEADHVILKANETF	479
Qy	481	AFVGNVTHYAQVWLNISAEIRSFLQGRLLQQLRWLQYVAELRLHPEALNLSDELPPA	540
		:	
Db	480	AFVGNVTHYAQVWLNISAEIRSFLQGRLLQQLRWLQYVADLRLHPEAMNLSDELPPA	539
Qy	541	LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQD	600

Db	540	LRLDYFSLPNGTALLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQD	599
Qy	601	NVTVFASVIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF	660
Db	600	NVTVFASVIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF	659
Qy	661	VWIQDMMERAIIDTFVGHADVVEPGSYVQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSV	720
Db	660	VWIQDMIERAIINTFVGHADVVEPGNYVQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSV	719
Qy	721	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAFITGFVQLSISVTALTAILKYGQVLMH	780
Db	720	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAFITGFVQLSISVTALTAILKYGQVLMH	779
Qy	781	SHVVIWLF LAVYAVATIMFCFLVSVLYSKAKLASACGGIIFLSYVPYMYVAIREEV AH	840
Db	780	SHVLIWLF LAVYAVATIMFCFLVSVLYSKAKLASACGGIIFLSYVPYMYVAIREEV AH	839
Qy	841	DKITAFEKCIASLMSTTAFGLGSKYFALYEAVGVGIQWHTFSQSPVEGDDFNLLLAVTML	900
Db	840	DKITAFEKCIASLMSTTAFGLGSKYFALYEAVGVGIQWHTFSQSPVEGDDFNLLLAVTML	899
Qy	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSV	960
Db	900	MVDTVVYGVL TWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTETWEWSWPWAHAPRLSV	959
Qy	961	MEEDQACAMESRRFEETRGMEEPTHLPVVCVDKLTKVYKDDKKLALNKLNLNLYENQV	1020
Db	960	MEEDQACAMESRHFEETRGMEEPTHLPVVCVDKLTKVYKNDKKLALNKLNLNLYENQV	1019
Qy	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR L	1080
Db	1020	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLF DQL	1079
Qy	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHS LVQTLSGGMKRKLSVAIAFVG	1140
Db	1080	TVEEHLWFYSRLKSMAQEEIRKEMDKMIEDLELSNKRHS LVQTLSGGMKRKLSVAIAFVG	1139
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Db	1140	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1199
Qy	1201	KCCGSPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Db	1200	KCCGSPLFLKGAYGDGYRLTLVKRPAEPGTSQEPGMASSPSGRPQLSNCSEMQVSQFIRK	1259
Qy	1261	HVASCLLVSDTSTELSYILPSEAAKKAFAERLFQHLERSLDALHLSSFGLMDTTLEEVFL	1320
Db	1260	HVASSLLVSDTSTELSYILPSEAVKKAFAERLFQQLHSLDALHLSSFGLMDTTLEEVFL	1319
Qy	1321	KVSEEDQSLENSEADV KESRKDVLPGAEGPASGEGHAGN LARCELTQSQASLQSASSVG	1380
Db	1320	KVSEEDQSLENSEADV KESRKDALPGAEGLTAVESQAGN LARCELAQSQASLQSASSVG	1379
Qy	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVQGGSRKLDGGWLKVRQF	1440



Db	1380	SARGDEGAGYTDGYGDYRPLFDNLQDPDSVSLQEAEMEALARVGQGSRKLEGWWLKMRQF	1439
Qy	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQ	1500
Db	1440	HGLLVKRFHCARRNSKALCSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQ	1499
Qy	1501	PRGNFIPIYANEERREYRLRLSPDASPQQVLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Db	1500	PRGNFIPIYANEERREYRLRLSPDASPQQVLVSTFRLPSGVGATCVLKSPANGSLGPMNLNS	1559
Qy	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDE-LQAWNVSPLPT	1619
Db	1560	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPLSPDEDSLLAWNTSLPPT	1619
Qy	1620	AGPEMWTSAAPSLPRLVREPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILTDTIGHNV	1679
Db	1620	AGPETWTWAPSLPRLVHEPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILTDTIGHNV	1679
Qy	1680	SEYLLFTSDRFLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYS	1739
Db	1680	SEYLLFTSDRFLHRYGAITFGNIQKSIPAPIGTRTPLMVRKIAVRRVAQVLYNNKGYS	1739
Qy	1740	MPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFI	1799
Db	1740	MPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFI	1799
Qy	1800	IVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWMLNYLVPATCCVIL	1859
Db	1800	IVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWMLNYLVPATCCIIIL	1859
Qy	1860	FVFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITA	1919
Db	1860	FVFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITA	1919
Qy	1920	TVATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDK	1979
Db	1920	TVATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEIAYNEYINEYYAKIGQFDK	1979
Qy	1980	MKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDQDVASERQ	2039
Db	1980	MKSPFEWDIVTRGLVAMTVEGFVGFLLTIMCQYNFLRQPQRLPVSTKPVEDDQDVASERQ	2039
Qy	2040	RVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLVNGAGKTSTFK	2099
Db	2040	RVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLVNGAGKTSTFK	2099
Qy	2100	MLTGDESTTGGEAFVNGHSVLKELLQVQVQSLGYCPQCDALFDELTAREHLQLYTRLRGIS	2159
Db	2100	MLTGDESTTGGEAFVNGHSVLKDLLQVQVQSLGYCPQFDALFDELTAREHLQLYTRLRGIP	2159
Qy	2160	WKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDP	2219
Db	2160	WKDEAQVVRWALEKLELTKCADKPAGSYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDP	2219
Qy	2220	KARRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGD	2279
Db	2220	KARRFLWNLILDLIKTGRSVVLTSHSMEECEAVCTRIAMVNGRLRCLGSIQHLKNRFGD	2279

Qy 2280 GYMITVRTKSSQSVKDVVRFFNRFPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVS 2339  
 |||||:|||||  
 Db 2280 GYMITVRTKSSQNVKDVVRFFNRFPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEHV 2339  
 Qy 2340 GVLGIEDYSVSQTTLDNVFNFAKKQSDNLEQQETEPPSALQSPLGCLLSLLRPRSAPTE 2399  
 |||||:|||||  
 Db 2340 GVLGIEDYSVSQTTLDNVFNFAKKQSDNVEQQEAE-PSTLPSPLG-LLSLLRPRPAPTE 2397  
 Qy 2400 LRALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2436  
 |||||  
 Db 2398 LRALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2434

RESULT 7

AAAY72649

ID AAY72649 standard; protein; 2001 AA.

XX

AC AAY72649;

XX

DT 31-MAY-2001 (first entry)

XX

DE Human ATP binding cassette2 (ABC2) transporter protein.

XX

KW Human; adenosine triphosphate; ATP; ATP binding cassette2 transporter;  
 KW ABC2 transporter; nootropic; neuroprotective; anticonvulsant; neurotoxic;  
 KW beta-amyloid; multidrug resistance; therapy; Alzheimer's disease;  
 KW prion disease; Parkinson's disease; Huntington's disease; panic disorder;  
 KW cholesterol misregulation; inflammatory disease; blood brain barrier;  
 KW cancer; mood disorder.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Misc-difference 30

FT /label= Unknown

FT /note= "Encoded by GSG"

FT Misc-difference 70

FT /label= Unknown

FT /note= "Encoded by TYC"

FT Domain 274. .296

FT /label= TMH

FT /note= "Transmembrane helix"

FT Misc-difference 280

FT /label= Unknown

FT /note= "Encoded by GYG"

FT Domain 317. .339

FT /label= TMH

FT /note= "Transmembrane helix"

FT Domain 351. .373

FT /label= TMH

FT /note= "Transmembrane helix"

FT Domain 380. .398

FT /label= TMH

FT /note= "Transmembrane helix"

FT Domain 411. .428

FT /label= TMH

FT /note= "Transmembrane helix"  
 FT Domain 457. .479  
 FT /label= TMH  
 FT /note= "Transmembrane helix"  
 FT Misc-difference 477  
 FT /label= Unknown  
 FT /note= "Encoded by MCG"  
 FT Misc-difference 558  
 FT /label= Unknown  
 FT /note= "Encoded by TKC"  
 FT Domain 588. .600  
 FT /label= Walker\_A  
 FT Region 689. .695  
 FT /label= ABC\_signature  
 FT Domain 696. .716  
 FT /label= Walker\_B  
 FT Domain 1022. .1044  
 FT /label= TMH  
 FT /note= "Transmembrane helix"  
 FT Domain 1358. .1380  
 FT /label= TMH  
 FT /note= "Transmembrane helix"  
 FT Domain 1410. .1432  
 FT /label= TMH  
 FT /note= "Transmembrane helix"  
 FT Domain 1441. .1463  
 FT /label= TMH  
 FT /note= "Transmembrane helix"  
 FT Domain 1470. .1492  
 FT /label= TMH  
 FT /note= "Transmembrane helix"  
 FT Misc-difference 1471  
 FT /label= Unknown  
 FT /note= "Encoded by GKG"  
 FT Domain 1553. .1575  
 FT /label= TMH  
 FT /note= "Transmembrane helix"  
 FT Domain 1650. .1662  
 FT /label= Walker\_A  
 FT Misc-difference 1651  
 FT /label= Unknown  
 FT /note= "Encoded by CYC"  
 FT Misc-difference 1689  
 FT /label= Unknown  
 FT /note= "Encoded by CHC"  
 FT Misc-difference 1720  
 FT /label= Unknown  
 FT /note= Encoded by CTN  
 FT Misc-difference 1724  
 FT /label= Unknown  
 FT /note= "Encoded by YCC"  
 FT Region 1751. .1758  
 FT /label= ABC\_signature  
 FT Domain 1759. .1780  
 FT /label= Walker\_B  
 XX  
 PN WO200114414-A2.

XX PD 01-MAR-2001.  
XX  
PF 18-AUG-2000; 2000WO-CA000962.  
XX  
PR 20-AUG-1999; 99US-0150073P.  
PR 30-AUG-1999; 99US-0151457P.  
PR 17-AUG-2000; 2000US-00641040.  
XX  
PA (ACTI-) ACTIVEPASS PHARM INC.  
XX  
PI Le Bihan S, Wilson C, Charest DL;  
XX  
DR WPI; 2001-202931/20.  
DR N-PSDB; AAD02722.  
XX  
PT Novel adenosine triphosphate (ATP) binding cassette transporter protein  
PT 2, useful as target for developing modulators that modulate activity of  
PT transporter protein and thus treat Alzheimer's disease and Parkinson's  
PT disease.  
XX  
PS Claim 13; Fig 2; 92pp; English.  
XX  
CC The present sequence is human adenosine triphosphate (ATP) binding  
CC cassette2 (ABC2) transporter protein. ABC2 transporter molecules are  
CC transmembrane proteins which catalyse ATP-dependent transport of  
CC endogenous or exogenous substrates across the biological membranes. ABC2  
CC transporters have been associated with the transport of neurotoxic  
CC polypeptides (e.g., beta-amyloid) and substrates across the blood-brain-  
CC barrier. ABC2 sequence is useful as target for developing modulators that  
CC are useful for modulating amyloid deposition and thus for treating  
CC Alzheimer's disease, prion diseases, Parkinson's disease and Huntington's  
CC disease. It is also useful as targets for developing modulating agents of  
CC multidrug resistance exhibited by e.g., cancer cells. ABC transporters  
CC are also useful for treating mood and panic disorders, cholesterol  
CC misregulation and inflammatory diseases. It can also be used to treat  
CC disorders characterised by insufficient or excessive production of an  
CC ABC2 transporter protein or its inhibitors. Fragments of ABC transporters  
CC are used as immunogens for producing antibodies  
XX  
SQ Sequence 2001 AA;

Qy	614	KDGS LPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGFVWIQDMMERAIID	673
Db	181	KDGS LPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGFVWIQDMMERAIID	240
Qy	674	TFVGHDVVEPGSYVQMFYPYCYTRDDFLVIEHMMPLCMVISWVYSVAMTIQHIVAEKEH	733
Db	241	TFVGHDVVEPGSYVQMFYPYCYTRDDFLVIEHMMPLCMXISWVYSVAMTIQHIVAEKEH	300
Qy	734	RLKEVMKTMGLNNAVHWVAWFITGFVQLSISVTALTALIKYGQVLMHSHVVIWFLAVY	793
Db	301	RLKEVMKTMGLNNAVHWVAWFITGFVQLSISVTALTALIKYGQVLMHSHVVIWFLAVY	360
Qy	794	AVATIMFCFLVSVLYSKAKLASACGGIIFYLSYVPYMYVAIREEVAHDKITAFЕКCІASL	853
Db	361	AVATIMFCFLVSVLYSKAKLASA-GGIIFYLSYVPYMYVAIREEVAHDKITAFЕКCІASL	419
Qy	854	MSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTMLMVDVAVYGILTWY	913
Db	420	MSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTMLMVDVAVYGILXWY	479
Qy	914	IEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSVMEEDQACAMESRR	973
Db	480	IEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSVMEEDQACAMESRR	539
Qy	974	FEETRGMEEEPTHLPLVVCVDKLTKVYKDDKKLALNKLSLNLYENQVVSFLGHNGAGKTT	1033
Db	540	FEETRGMEEEPTHLPLVXVDKLTKVYKDDKKLALNKLSLNLYENQGVVSFLGHNGAGKTT	599
Qy	1034	TMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSRLK	1093
Db	600	TMSILTGLFPPTSGSATIYGHDIRTEMDEIRKN-GHVPQHNVLFDRLTVEEHLWFYSRLK	658
Qy	1094	SMAQEEIRREMDKMIEDLELSNKRHSIVQTLSGGMKRKLSVAIAFVGGSRAIILDEPTAG	1153
Db	659	SMAQEEIPREMDKMIEDLELSNKRHSIVQTLSGGMKRKVSVAIAFVGGSRAIILDEPTAG	718
Qy	1154	VDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPLFLKGTY	1213
Db	719	VDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPLFLKGTY	778
Qy	1214	GDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHHVASCLLVSDTST	1273
Db	779	GDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHHVASCLLVSDTST	838
Qy	1274	ELSYILPSEAAKKGA FERLFQHLERSLDALHLSSFGIMDTTLEEVLKVSEEDQSLENS	1333
Db	839	ELSYILPSEAAKKGA FERLFQHLERSLDALHLSSFGIMDTTLEEVLKVS GGDQSLENSG	898
Qy	1334	ADV KESRKDVLP GAEGPASGEHAGN LARCSELTQS QASLQSASSVGSARGDEGAGYTDV	1393
Db	899	ADV KESRKDVLP GAEGHASGEHAGN LARCSELTQS QASLQSASSVGSALGDEGAGYTDV	958
Qy	1394	YGDYRPLFDNPQDPDNVSLQEVEAEALS RVGQGSRKLDGGWLKVRQFHGLLVKRFHCARR	1453
Db	959	YGDYPPLFDNPQDPDNVSLQEVEAEALS RVGQGSRKLDGGWLKVRQFHGLLVKRFHCARR	1018

Qy	1454	NSKALFSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYHNYTQPRGNFIPYANEER	1513
Db	1019	NSKALFSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYHNYTQPRGNFIPYANEER	1078
Qy	1514	REYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFD	1573
Db	1079	REYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFD	1138
Qy	1574	SMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS LPPTAGPEMWT SAPSLPR	1633
Db	1139	SMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS LPPTAGPEMWT SAPSLPR	1198
Qy	1634	LVREPVRTC CSAQGTGFSCPSSVGGHPPQMRVVTGDILTDITGHNVS EYLLFTSDRFR LH	1693
		:	
Db	1199	LVREPVRTC CSAQGTGFSCPNSVGGHPPQMRVVTGDILTDITGHNVS EYLLFTSDRFR LH	1258
Qy	1694	RYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSMPTYLNSLNNAILR	1753
Db	1259	RYGAITFGNVLKSIPASFGTRAPPMVRKIRCARAAQVFYNNKGYHSMPTYLNSLNNAILR	1318
Qy	1754	ANLPKSKGNPAAYGITVTNHMPNKT SASLSLDYLLQGT DVVIAIFII VAMSFVPASFVVF	1813
Db	1319	ANLPKSKGNPAAYGITVTNHMPNKT SASLSLDYLLQGT DVVIAIFII VAMSFVPASFVVF	1378
Qy	1814	LVAEKSTKAKHLQFVSGCNPIIYWLANYVWMLNYLVPATCCV IILFVFDLPAYTSPTNF	1873
Db	1379	LVAEKSTKAKHLQFVSGCNPIIYWLANYVWMLNYLVPATCCV IILFVFDLPAYTSPTNF	1438
Qy	1874	PAVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGITATVATFLLQLFEHDK	1933
Db	1439	PAVLSLFLLYGWSITPIMYPASFWFEVPSSAYXFLIVINLFIGITATVATFLLQLFEHDK	1498
Qy	1934	DLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDMKSPFEWDIVTRGL	1993
Db	1499	DLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDMKSPFEWDIVTRGL	1558
Qy	1994	VAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDD VDVASERQ RVLRGDADNDMVKI	2053
Db	1559	VAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDD VDVASERQ RVLRGDADNDMVKI	1618
Qy	2054	ENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLG VNGAGKTSTFKMLTGDESTTGGEAF	2113
Db	1619	ENLTKVYKSRKIGRILAVDRLCLGVRPGECFGXLGVNGAGKTSTFKMLTGDESTTGGEAF	1678
Qy	2114	VNGHSV LKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEK	2173
Db	1679	VNGHSV LKELXQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGIXWKDEARVVKWALEK	1738
Qy	2174	LELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLI LDI	2233
Db	1739	LELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLI LDI	1798
Qy	2234	KTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVVRTKSSQSV	2293
Db	1799	KTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVVRTKSSQSV	1858
Qy	2294	KDVVRFFNRNFP EAM LKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTT	2353

```

      |||
Db      1859 KDVVRFFNRFPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTT 1918
      |||
Qy      2354 LDNVFVNFQAKKQSDNLEQQETEPPSALQSPLGCLLSLLRPRSAPTELRLVADEPEDLDT 2413
      |||
Db      1919 LDNVFVNFQAKKQSDNLEQQETEPPSALQSPLGCLLSLLRPRSAPTELRLVADEPEDLDT 1978
      |||
Qy      2414 EDEGLISFEEERAQLSFNTDTLC 2436
      |||
Db      1979 EDEGLISFEEERAQLSFNTDTLC 2001

```

RESULT 8

ABB98347

ID ABB98347 standard; protein; 2001 AA.

XX

AC ABB98347;

XX

DT 29-JAN-2003 (first entry)

XX

DE Human ABC transporter ABCA2 SEQ ID NO 8.

XX

KW Human; ABC transporter; ABCB9; ABCB1; ABCA2; ABCG4; ABCG1;  
 KW amyloid precursor protein; adenosine tri-phosphate; nootropic;  
 KW ATP-binding cassette transporter; beta-amyloid plaque formation;  
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease;  
 KW gene therapy; transgenic; neuroprotective; anticonvulsant;  
 KW antiparkinsonian.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Misc-difference 30

FT /note= "Encoded by GCG"

FT Misc-difference 70

FT /note= "Encoded by TTC"

FT Misc-difference 92

FT /note= "Encoded by CTG"

FT Misc-difference 132

FT /note= "Encoded by GCC"

FT Misc-difference 174

FT /note= "Encoded by AGT"

FT Misc-difference 280

FT /note= "Encoded by GTG"

FT Misc-difference 383. .384

FT /note= "Encoded by GCCTGCGGT"

FT Misc-difference 477

FT /note= "Encoded by ACG"

FT Misc-difference 558

FT /note= "Encoded by TGC"

FT Misc-difference 586

FT /note= "Encoded by GTG"

FT Misc-difference 632. .635

FT /note= "Encoded by AACCTGGGCATGTGC"

FT Misc-difference 666

FT /note= "Encoded by CGC"

FT Misc-difference 697

FT /note= "Encoded by CTG"  
 FT Misc-difference 889  
 FT /note= "Encoded by GAG"  
 FT Misc-difference 890  
 FT /note= "Encoded by GAG"  
 FT Misc-difference 898  
 FT /note= "Encoded by GAG"  
 FT Misc-difference 915  
 FT /note= "Encoded by CCG"  
 FT Misc-difference 948  
 FT /note= "Encoded by CGT"  
 FT Misc-difference 963  
 FT /note= "Encoded by CGC"  
 FT Misc-difference 1187  
 FT /note= "Encoded by CCA"  
 FT Misc-difference 1219  
 FT /note= "Encoded by AGC"  
 FT Misc-difference 1288  
 FT /note= "Encoded by GCG"  
 FT Misc-difference 1289  
 FT /note= "Encoded by GTG"  
 FT Misc-difference 1290  
 FT /note= "Encoded by CGC"  
 FT Misc-difference 1471  
 FT /note= "Encoded by GTG"  
 FT Misc-difference 1651  
 FT /note= "Encoded by CTC"  
 FT Misc-difference 1689  
 FT /note= "Encoded by CTC"  
 FT Misc-difference 1724  
 FT /note= "Encoded by TCC"  
 XX  
 PN WO200264781-A2.  
 XX  
 PD 22-AUG-2002.  
 XX  
 PF 08-FEB-2002; 2002WO-CA000138.  
 XX  
 PR 09-FEB-2001; 2001US-0267975P.  
 PR 31-JUL-2001; 2001US-0309256P.  
 XX  
 PA (ACTI-) ACTIVE PASS PHARM INC.  
 XX  
 PI Reiner PB, Connop BP, Pollard M;  
 XX  
 DR WPI; 2002-667006/71.  
 DR N-PSDB; ABV74350.  
 XX  
 PT Regulating expression of amyloid precursor protein in a cell, useful in  
 PT preventing or treating neurological disease, e.g. Alzheimer's disease,  
 PT comprises regulating the expression or activity of an ATP-binding  
 PT cassette transporter.  
 XX  
 PS Disclosure; Page; 78pp + Sequence Listing; English.  
 XX  
 CC The invention relates to regulating (M1) expression of amyloid precursor  
 CC protein in a cell, comprising regulating the expression or activity of an



CC adenosine tri-phosphate (ATP)-binding cassette (ABC) transporter in the  
 CC cell. (M1) is useful for regulating expression of amyloid precursor  
 CC protein in a brain cell to prevent or inhibit pathological beta-amyloid  
 CC plaque formation in conditions such as Alzheimer's disease, Parkinson's  
 CC disease or Huntington's disease. (M1) is also useful in screening assays,  
 CC predictive medicine (e.g. diagnostic assays, prognostic assays,  
 CC monitoring clinical trials or phamacogenetics) or methods of treatment  
 CC (e.g. therapeutic, prophylactic, gene therapy). The transgenic animals  
 CC are useful for testing methods and agents as candidates for modulating or  
 CC altering the ABC transporter-relates expression of amyloid precursor  
 CC protein. The present sequence is that of an ABC transporter protein  
 CC encoding polynucleotide of the invention. Note: The sequence data for  
 CC this patent is not represented in the printed specification but is based  
 CC on sequence information supplied to Derwent by the European Patent Office  
 XX

SQ Sequence 2001 AA;

Query Match 80.9%; Score 10249; DB 5; Length 2001;  
 Best Local Similarity 98.5%; Pred. No. 0;  
 Matches 1973; Conservative 2; Mismatches 26; Indels 2; Gaps 2;

Qy	434	MSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVW	493
Db	1	MSSLGFTSKEQRNLGLLVHLMTSNPKILYXPAGSEVDRVILKANETFAFVGNVTHYAQVW	60
Qy	494	LNISAEIRSFLQGRLLQQLRLRWLQQYVAELRLHPEALNLSLDELPPALRQDNFSLPSGMA	553
Db	61	LNISAEIRSXLEQGRLLQQLRLRWLQQYVAELRPHPEALNLSLDELPPALRQDNFSLPSGMA	120
Qy	554	LLQQLDITIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQDNVTVFASVIFQTR	613
Db	121	LLQQLDITIDNAPCGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQDNVTVFAGVIFQTR	180
Qy	614	KDGSLLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGFVWIQDMMERAIID	673
Db	181	KDGSLLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGFVWIQDMMERAIID	240
Qy	674	TFVGHDDVVEPGSYVQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSVAMTIQHIVAEKEH	733
Db	241	TFVGHDDVVEPGSYVQMFYPCYTRDDFLFVIEHMMPLCMXISWVYSVAMTIQHIVAEKEH	300
Qy	734	RLKEVMKTMGLNNAVHVAWFITGFVQLSISVTALTAILKYGQVLMHSHVVIWLFLAVY	793
Db	301	RLKEVMKTMGLNNAVHVAWFITGFVQLSISVTALTAILKYGQVLMHSHVVIWLFLAVY	360
Qy	794	AVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHDKITAFEKCIASL	853
Db	361	AVATIMFCFLVSVLYSKAKLASA-GGIIYFLSYVPYMYVAIREEVAHDKITAFEKCIASL	419
Qy	854	MSTTAFGLGSKYFALYEAVGVGIQWHTFSQSPVEGDDFNLLAVTMLMVDVAVYGILTWY	913
Db	420	MSTTAFGLGSKYFALYEAVGVGIQWHTFSQSPVEGDDFNLLAVTMLMVDVAVYGILXWY	479
Qy	914	IEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSVMEEDQACAMESRR	973
Db	480	IEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSVMEEDQACAMESRR	539

Qy	974	FEETRGMEEETHLPLVVCVDKLTKVYKDDKKLALNKLNLNLYENQVVSFLGHNGAGKTT	1033
Db	540	FEETRGMEEETHLPLVVXVDKLTKVYKDDKKLALNKLNLNLYENQGVVSFLGHNGAGKTT	599
Qy	1034	TMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSRLK	1093
Db	600	TMSILTGLFPPTSGSATIYGHDIRTEMDEIRKN-GHVPQHNVLFDRLTVEEHLWFYSRLK	658
Qy	1094	SMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVGGSRAILDEPTAG	1153
Db	659	SMAQEEIPREMDKMIEDLELSNKRHSLVQTLSSGGMKRKVSVAIAFVGGSRAILDEPTAG	718
Qy	1154	VDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPLFLKGTY	1213
Db	719	VDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPLFLKGTY	778
Qy	1214	GDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHAVASCLLVSDTST	1273
Db	779	GDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHAVASCLLVSDTST	838
Qy	1274	ELSYILPSEAAKKGAFERLFQHLERSLDALHLSSFGLMDTTLEEVFLKVSEEDQSLENSE	1333
Db	839	ELSYILPSEAAKKGAFERLFQHLERSLDALHLSSFGLMDTTLEEVFLKVSAGDQSLNSG	898
Qy	1334	ADVKEsrkdVLPgaEGPASGEGHAGNLARCSELTQSQASLQSASSVGSARGDEGAGYTDV	1393
Db	899	ADVKEsrkdVLPgaEGHASGEGHAGNLARCSELTQSQASLQSASSVGSALGDEGAGYTDV	958
Qy	1394	YGDYRPLFDNPQDPDNVSLQEEAEALSRVQGSRKLDGGWLKVRQFHGLLVKRFHCARR	1453
Db	959	YGDYPPLFDNPQDPDNVSLQEEAEALSRVQGSRKLDGGWLKVRQFHGLLVKRFHCARR	1018
Qy	1454	NSKALFSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYHNYTQPRGNFIPIYANEER	1513
Db	1019	NSKALFSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYHNYTQPRGNFIPIYANEER	1078
Qy	1514	REYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFD	1573
Db	1079	REYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFD	1138
Qy	1574	SMCLESFTQGLPLSNFVPPPPSPAPSDSPASPEDLQAWNVS LPPTAGPEMWTSA PSLPR	1633
Db	1139	SMCLESFTQGLPLSNFVPPPPSPAPSDSPASPEDLQAWNVS LPPTAGQEMWTSA PSLPR	1198
Qy	1634	LVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTDTIGHNVSEYLLFTSDRFR LH	1693
Db	1199	LVREPVRCTCSAQGTGFSCPNSVGGHPPQMRVVTGDILTDTIGHNVSEYLLFTSDRFR LH	1258
Qy	1694	RYGAIITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSMPTYLNSLNNA ILR	1753
Db	1259	RYGAIITFGNVLKSIPASFGTRAPPMVRKIRCARAAQVFYNNKGYHSMPTYLNSLNNA ILR	1318
Qy	1754	ANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDDVIAIFIIVAMS FVPASFVVF	1813
Db	1319	ANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDDVIAIFIIVAMS FVPASFVVF	1378
Qy	1814	LVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNVLVPATCCVILFVFDLPAYTSPTNF	1873

Db	1379	LVAEKSTKAKHLQFVSGCNPIIYWLANYVWMLNLYLVPATCCVILFVFDLPAYTSPTNF	1438
Qy	1874	PAVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITATVATFLLQLFEHDK	1933
Db	1439	PAVLSLFLLYGWSITPIMYPASFWEVPSSAYXFLIVINLFIGITATVATFLLQLFEHDK	1498
Qy	1934	DLKVVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGL	1993
Db	1499	DLKVVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGL	1558
Qy	1994	VAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDQDVASERQVRVLRGDADNDMVKI	2053
Db	1559	VAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDQDVASERQVRVLRGDADNDMVKI	1618
Qy	2054	ENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLG VNGAGKTSTFKMLTGDESTTGGEAF	2113
Db	1619	ENLTKVYKSRKIGRILAVDRLCLGVRPGECFGX LGVNGAGKTSTFKMLTGDESTTGGEAF	1678
Qy	2114	VNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEK	2173
Db	1679	VNGHSVLKELXQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGIXWKDEARVVKWALEK	1738
Qy	2174	LELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLI	2233
Db	1739	LELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLI	1798
Qy	2234	KTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVRTKSSQSV	2293
Db	1799	KTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVRTKSSQSV	1858
Qy	2294	KDVVRFFNRFPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTT	2353
Db	1859	KDVVRFFNRFPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTT	1918
Qy	2354	LDNVFVNFAKKQSDNLEQQETEPSPALQSPLGCLLSLLRPRSAPTELRALVADEPEDLDT	2413
Db	1919	LDNVFVNFAKKQSDNLEQQETEPSPALQSPLGCLLSLLRPRSAPTELRALVADEPEDLDT	1978
Qy	2414	EDEGLISFEEERAQLSFNTDTLC	2436
Db	1979	EDEGLISFEEERAQLSFNTDTLC	2001

RESULT 9

AAE16781

ID AAE16781 standard; protein; 1771 AA.

XX

AC AAE16781;

XX

DT 09-APR-2002 (first entry)

XX

DE Human transporter and ion channel-18 (TRICH-18) protein.

XX

KW Human; transporter and ion channel-18; TRICH-18; neuroprotective; asthma;  
 KW nootropic; cytostatic; cardiovascular; immunosuppressive; cardiomyopathy;  
 KW antiinflammatory; protein therapy; akinesia; cystic fibrosis; leukaemia;

KW Bell's palsy; amyotrophic lateral sclerosis; Alzheimer's disease; cancer;  
 KW amnesia; dementia; myocarditis; Duchenne's muscular dystrophy; AIDS;  
 KW Acquired Immune Deficiency Syndrome; Addison's disease; allergy; angina;  
 KW cell proliferative disorder; psoriasis; cardiac disease; hypertension;  
 KW bradyarrhythmia; gene expression; drug screening.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 119. .138  
 FT /label= Transmembrane\_domain  
 FT Domain 228. .246  
 FT /label= Transmembrane\_domain  
 FT Domain 353. .533  
 FT /note= "ABC transporter domain"  
 FT Binding-site 360. .367  
 FT /label= P\_loop  
 FT /note= "ATP/GTP binding site"  
 FT Domain 1128. .1148  
 FT /label= Transmembrane\_domain  
 FT Domain 1180. .1197  
 FT /label= Transmembrane\_domain  
 FT Domain 1235. .1261  
 FT /label= Transmembrane\_domain  
 FT Domain 1416. .1597  
 FT /note= "ABC transporter domain"  
 FT Binding-site 1423. .1430  
 FT /label= P\_loop  
 FT /note= "ATP/GTP binding site"  
 XX  
 PN WO200192304-A2.  
 XX  
 PD 06-DEC-2001.  
 XX  
 PF 25-MAY-2001; 2001WO-US017065.  
 XX  
 PR 26-MAY-2000; 2000US-0208424P.  
 PR 01-JUN-2000; 2000US-0209001P.  
 PR 08-JUN-2000; 2000US-0210588P.  
 PR 16-JUN-2000; 2000US-0212335P.  
 PR 22-JUN-2000; 2000US-0213747P.  
 PR 29-JUN-2000; 2000US-0215391P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Thornton M, Walia NK, Yue H, Nguyen DB, Lai P, Gandhi AR;  
 PI Tribouley CM, Yao MG, Ramkumar J, Au-Young J, Lu Y, Tang YT;  
 PI Azimzai Y, Bruns CM, Griffin JA, Yang J, Sanjanwala MS, Raumann BE;  
 PI Lee EA, Hafalia A, Baughn MR, Green BD, Khan FA, Kearney L;  
 PI Elliot VS, Seilhamer JJ, Policky JL, Borowsky ML, Burford N, Ding L;  
 PI Lu DAM, Hillman JL;  
 XX  
 DR WPI; 2002-122055/16.  
 DR N-PSDB; AAD27271.  
 XX  
 PT New human transporters and ion channels (TRICH) polypeptides useful for  
 PT diagnosing, treating or preventing disorders associated with aberrant



Db	421	LWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVGGSGRAI	480
Qy	1146	ILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGS	1205
Db	481	ILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGS	540
Qy	1206	PLFLKGTYGDDYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHVASC	1265
Db	541	PLFLKGTYGDDYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHVASC	600
Qy	1266	LLVSDTSTELSYILPSEAAKKGAFERLFQHLESLDALHLSSFGMLDTTLEEVFLKVSEE	1325
Db	601	LLVSDTSTELSYILPSEAAKKGAFERLFQHLESLDALHLSSFGMLDTTLEEVFLKVSEE	660
Qy	1326	DQSELENSEADVKEARKDVLPGAEGPASGEGHAGNLARCSELTQSQASLQSASSVSGARGD	1385
Db	661	DQSELENSEADVKEARKDVLPGAEGPASGEGHAGNLARCSELTQSQASLQSASSVSGARGD	720
Qy	1386	EGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQSGSRKLDGGWLKVRQFHGLLV	1445
Db	721	EGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQSGSRKLDGGWLKVRQFHGLLV	780
Qy	1446	KRFHCARRNSKALFSQILLPAFFVCVAMTVALSVP EIGDLPPLVLSPSQYHNYTQPRGNF	1505
Db	781	KRFHCARRNSKALFSQILLPAFFVCVAMTVALSVP EIGDLPPLVLSPSQYHNYTQPRGNF	840
Qy	1506	IPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGESR	1565
Db	841	IPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGESR	900
Qy	1566	LLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS LPPTAGPEMW	1625
Db	901	LLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS LPPTAGPEMW	960
Qy	1626	TSAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILT DITGHNVS EYLLF	1685
Db	961	TSAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILT DITGHNVS EYLLF	1020
Qy	1686	TSDRFR L HRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKG YHSMPTYLN	1745
Db	1021	TSDRFR L HRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKG YHSMPTYLN	1080
Qy	1746	SLNNAILRANLPKSKGNPAAYGITVTNHMPMNKTSASLSLDYLLQGT DVVIAIFIIVAMSF	1805
Db	1081	SLNNAILRANLPKSKGNPAAYGITVTNHMPMNKTSASLSLDYLLQGT DVVIAIFIIVAMSF	1140
Qy	1806	VPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDM LNYLVPATCCV IILFVFDLP	1865
Db	1141	VPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDM LNYLVPATCCV IILFVFDLP	1200
Qy	1866	AYTSPTNFP AVL S LFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITATVATFL	1925
Db	1201	AYTSPTNFP AVL S LFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITATVATFL	1260
Qy	1926	LQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFE	1985
Db	1261	LQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFE	1320

Qy 1986 WDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDDVDVASERQRVLRGD 2045  
 |||  
 Db 1321 WDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDDVDVASERQRVLRGD 1380  
 Qy 2046 ADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLGVNAGKTSTFKMLTGDE 2105  
 |||  
 Db 1381 ADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLGVNAGKTSTFKMLTGDE 1440  
 Qy 2106 STTGGEAFVNGHSLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEAR 2165  
 |||  
 Db 1441 STTGGEAFVNGHSLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEAR 1500  
 Qy 2166 VVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFL 2225  
 |||  
 Db 1501 VVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFL 1560  
 Qy 2226 WNLILDLIKTRGSVVLTSMSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITV 2285  
 |||  
 Db 1561 WNLILDLIKTRGSVVLTSMSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITV 1620  
 Qy 2286 RTKSSQSVKDVVRFFNRNFPPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIE 2345  
 |||  
 Db 1621 RTKSSQSVKDVVRFFNRNFPPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIE 1680  
 Qy 2346 DYSVSQTTLDNVFVNFQKQSDNLEQQETEPSPALQSPLGCLLSLLRPRSAPTELRAVA 2405  
 |||  
 Db 1681 DYSVSQTTLDNVFVNFQKQSDNLEQQETEPSPALQSPLGCLLSLLRPRSAPTELRAVA 1740  
 Qy 2406 DEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2436  
 |||  
 Db 1741 DEPEDLDTEDEGLISFEEERAQLSFNTDTLC 1771

RESULT 10

AAB38110

ID AAB38110 standard; protein; 2261 AA.

XX

AC AAB38110;

XX

DT 29-JAN-2001 (first entry)

XX

DE Human ABC1 cholesterol transporter mutant, V399A.

XX

KW Human ABC1 cholesterol transporter; chromosome 9q31;

KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;

KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;

KW cardiovascular disease; coronary artery disease; coronary restenosis;

KW cerebrovascular disease; peripheral vascular disease;

KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;

KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;

KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;

KW mutein.

XX

OS Homo sapiens.

XX

PN WO200055318-A2.

XX  
PD 21-SEP-2000.  
XX  
PF 15-MAR-2000; 2000WO-IB000532.  
XX  
PR 15-MAR-1999; 99US-0124702P.  
PR 08-JUN-1999; 99US-0138048P.  
PR 17-JUN-1999; 99US-0139600P.  
PR 01-SEP-1999; 99US-0151977P.  
XX  
PA (UYBR-) UNIV BRITISH COLUMBIA.  
PA (XENO-) XENON BIORESEARCH INC.  
XX  
PI Hayden MR, Wilson AR, Pimstone SN;  
XX  
DR WPI; 2000-587528/55.  
XX  
PT New ABC1 polypeptide is useful for treating diseases associated with ABC1  
PT biological activity, e.g. Alzheimer's disease, Huntington's disease and  
PT cancer.  
XX  
PS Example; Page; 229pp; English.  
XX  
CC The invention relates to the human ABC1 cholesterol transporter protein  
CC (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is  
CC a member of the ATP-binding cassette (ABC transporter) superfamily of  
CC proteins, and plays a crucial role in cholesterol transport, particularly  
CC intracellular cholesterol trafficking in monocytes and fibroblasts, being  
CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is  
CC located on chromosome 9q31, and mutations in this gene are associated  
CC with two genetic HDL (high density lipoprotein) deficiency disorders,  
CC Tangier disease (TD) and familial HDL deficiency (FHA). These diseases  
CC are distinguishable in that TD is an autosomal recessive disorder, while  
CC FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good  
CC cholesterol") in the blood correlate with a high risk of cardiovascular  
CC disease, particularly coronary artery disease, but also cerebrovascular  
CC disease, coronary restenosis, and peripheral vascular disease.  
CC Conversely, a high level of HDL has protective effects against  
CC cardiovascular disease. The invention provides genetic constructs and  
CC transgenic cells and non-human animals comprising human ABC1 nucleic  
CC acids, and methods of gene therapy for the treatment or prevention of  
CC cardiovascular disease comprising the administration of an expression  
CC vector encoding ABC1 or an active fragment thereof. The invention also  
CC encompasses compounds which mimic ABC1 activity, compounds which  
CC stimulate ABC1 expression and methods of screening for such compounds. It  
CC further relates to methods for determining whether a patient has an  
CC increased risk for cardiovascular disease due to polymorphisms in the  
CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat or  
CC prevent cardiovascular disease, especially coronary artery disease,  
CC cerebrovascular disease, coronary restenosis or peripheral vascular  
CC disease. They may also be used in the treatment of diseases associated  
CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick  
CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.  
CC The invention specifically excludes proteins with the exact amino acid  
CC sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic  
CC acid with the exact sequence as GenBank Accession No: AJ012376.1. The  
CC present sequence represents a mutant human ABC1 cholesterol transporter



CC associated with an altered cholesterol level and therefore an altered  
CC risk of cardiovascular disease. Note: The present sequence is not shown  
CC in the specification, but is derived from the native human ABC1 shown on  
CC pages 152-157

XX

SQ Sequence 2261 AA;

Query Match 33.5%; Score 4244.5; DB 3; Length 2261;  
Best Local Similarity 39.9%; Pred. No. 1.2e-307;  
Matches 1001; Conservative 345; Mismatches 729; Indels 435; Gaps 61;

```
Qy      6 QLQLLLWKNVTLKRRSPWVLAFEIFIPVLVFFILLGLRQKKPTISVKEVPFYTAAPL TSA 65
      ||:|||||:| :|| | | :| ||: :| | | :| |
Db      6 QLRLLLWKNLTFRRRQTCQLLLEVAWPLFIFLILISVRLSYPPYEQHECHFNPKA-MPSA 64

Qy     66 GILPVMQSLCPDGQRDEFGFL-----QYANSTVTQLLERLDRVVEEGNLFDPARP 115
      | || :| : : | : : | | :| | :| :|
Db     65 GTLPVWQGIICNANNPCFRYPTPGEAPGVVGNFNKSIVARLFSDARRLL----LYSQKDT 120

Qy    116 SLGSELEALR--QHLEALSAGPGTSGSHLDRSTVSSFSLDVARNPQELWRFLTQNL SLP 173
      |: : || | : : | : : | | | | | | | |
Db    121 SMKDMRKVLRTLQQIKSSSNLKLQDFLVDNETFSG-----FLYHNLSLP 165

Qy    174 NSTAQALLAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELL LPA 233
      || :| | | :| :| | | | | :| | :|
Db    166 KSTVDKMLRADV---ILHKVFLQGYQLHLTS-LCNGS-----KSEEMI---- 204

Qy    234 LLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAELRNQLDV 293
      || | | :| : :| :| : | | | | | | | |
Db    205 ---QL---GDQEVSELCGLPREKLAAAE-----RVLRSNMDI 235

Qy    294 AK-VSQQGLGLDAPNGSDSSPQAPPPRRQLQALLGDL LD-----AQKVLQDVDVLS 341
      | : : | :| | :| :| | | | | | | | |
Db    236 LKPILRTLNSTSPFPSKELAEA--TKTLHSLGTLAQELFSMRSWSDMRQEV MFLT NVNS 293

Qy    342 ALALLLPQGACTGRTPGPPASGAGGAAN-----GTGAGAVMGP NATAE EGAPSAAALATP 396
      : : | : | | | : : | | | | | : | |
Db    294 SSSSTQIYQAVSRIVCGHPEGGLKIKSLNWYEDNNYKALFGGNGTEEDAETFYDNSTTP 353

Qy    397 ---DTLQGQCSAFVQ--LWAGLQPILCGNNRTIEPEALRRGNMSSLGFTSKEQRNLGLLV 451
      | : : | : : :| | :| :| | |
Db    354 YCNDLMKNLESSPLSRIIWKALKPLLVG----- 381

Qy    452 HLMTSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVWLNISAEIRS FLEQGR LQQ 511
      |||| | :| :| :| :| : | :| :| :| :| :|
Db    382 -----KILYTPDTPATRQVMAEANKTFQELAVFHDLEGMWHEELSPKIWTFMENSQEMD 434

Qy    512 HLRWL-----QQYVAELRLHPE---ALNLSLDELPPALRQDNFS 547
      :| | | | | | | | | | | | :| | :|
Db    435 LVRMLLDSRDNDHFWEQQLDGLDWT AQDIVAFLAKHPEDVQSSNGSVYT WREAFNETN-- 492

Qy    548 LPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPEESIVNYTLNQAYQDNVTVFAS 607
      | : || :|| | : : : | :| :| :| :|
Db    493 -----QAIRTIS-----RFMECVNLNLEPIATEVWLINKSME--LLDERKFWAG 535

Qy    608 VIFQTRKDGS--LPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTG---GRFYFLYGFVW 662
      ::| || || || ||| : | :| :| :| || || | || :
```

Db 536 IVFTGITPGSIELPHHVYKIRMDIDNVERTNKKIDGYWDPGPRADPFEDMRYVWGGFAY 595

Qy 663 IQDMMERAIIDTFVGHDVVEPGSYVQMFPYPCYTRDDFLFVIEHMMPLCMVISWVYSVAM 722  
:|:|:|:|:| | : : | |:| | | | | | : | | | :|:|:|:|:|:

Db 596 LQDVVEQAIIRVLTGTE-KKTGVYMQQMPYPCYVDDIFLRVMSRSMPLFMTLAWIYSVAV 654

Qy 723 TIQHIVAEKEHRLKEVMKTMGLNNAVHWVAVFITGFVQLSISVTALTALIKYGQVIMHSH 782  
|: | | | | | | | | : | | | :|:|:|:|:| : | :| | | | :|:|

Db 655 IIKGIVYEKEARLKETMRIMGLDNSILWFSWFISSLIPLLVSAGLLVVILKLGNNLLPYSD 714

Qy 783 VVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHDK 842  
:|:|:|:|:| | | : | | :| :|:| | | | | | | | | | | | | | | | |

Db 715 PSVVFVFLSVFAVVTILQCFLISTLFSRANLAAACGGIIYFTLYLPYVLC-----VAWQD 769

Qy 843 ITAFE-KCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTMLM 901  
| | | | | | | | | | | : | | | :|:|:|:|:| | | | | | | :|:|:|:

Db 770 YVGFTLKFIFASLLSPVAFGFGCEYFALFEEQGIGVQWDLNLFESPVEEDGFNLTTSVSMML 829

Qy 902 VDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSVM 961  
| :|:|:|:|:| | | | | | | | | | | | | | | | | : | | | : |:|

Db 830 FDTFLYGVMTWYIEAVFPGQYGI PRPWYFPCTKSYWGE---ESDEKSHPGSNQKRIS-- 884

Qy 962 EEDQACAMESRRFEETRGMEEEPTHLPLVVCVDKLTkVYKDDKKLALNKLNLNLYENQVV 1021  
: | | | | | | | | | | : | | | :|:|:|:|:| | | | | | |

Db 885 ---EIC-----MEEPTHLKLGVSIGNLVKVYRDGMKVAVDGLALNFYEGQIT 929

Qy 1022 SFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLT 1081  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 930 SFLGHNGAGKTTTMSILTGLFPPTSGTAYILGKDIREMSTIRQNLGVCPQHNVLFDMLT 989

Qy 1082 VEEHLWFYSRLKSMAQEEIRREMDKMIEDLEL-SNKRHSIVQTLSSGGMKRKLSVAIAFVG 1140  
| | | :|:|:|:| | :|:| :|:| | : | | :| | | | | | | | | | | | |

Db 990 VEEHIWFYARLKGLSEKHVKAEMEQLMDVGLPSSKLKSKTSQLSGGMQRKLSVALAFVG 1049

Qy 1141 GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL 1200  
| | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 1050 GSKVVILDEPTAGVDPYSRRGIWELLLKYRQGRTIILSTHHMDEADVLDRIAIISHGKL 1109

Qy 1201 KCCGSPLFLKGTYG DGYRLTLVKRPAEPG-----GPQEPGLAS 1238  
| | | | | | | | | | | | | | : | | | | | | : | | |

Db 1110 CCVGSSLFLKNQLGTGYLTLVKKDVESSLSSCRNSSSTVSYLKKEDSVSQSSSDAGLGS 1169

Qy 1239 SPPGRAPLSSCSELQVSQFIRKHVASCLIVSDTSTELSYILPSEAAKKGA FERL FQHLE 1298  
| : | | | | : | | | | | | | | :|:|:|:| | | | | | | | | |

Db 1170 DHESDTLTIDVS--AISNLIRKHVSEARLVEDIGHELTYVLPYEAKEGAFVELFHEIDD 1227

Qy 1299 SLDALHLSSFGLMDTTLEEVFLKVSEEDQSLSENSEADV KESRKDVLPGAEGPASGEGHAG 1358  
| | :|:|:|:| :| | | | | | | | | | | | | | | | | | | | | |

Db 1228 RLSDLGISSYGISSETTLEEIFLKVAEE-----SGVDA-ETSDGTLP----- 1267

Qy 1359 NLARCELTQSQASLQSASSVGSARGDEGAGYTDVYGDYRPLF-DNPQDPD--NVSLQEV 1415  
| | : | : | | | | | : | | :|:|:|:|

Db 1268 -----ARRNRA-FGDKQSCLRPFTEDDAADPNDSIDIPESR 1303

Qy 1416 EAEALSRV-GQGSRKLDGGWLKVRQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMT 1474  
| : | | : |:| | :| | | :| | | | | | | | :|:|:|:| | | | |

Db 1304 ETDLLSGMDGKGSYQVGWKLTTQQQFVALLWKRLLIARRSRKGFFAQIVLPAVFVCIALV 1363

Qy 1475 VALSVPEIGDLPLVLSPSQYH-NYTQPRGNFIPIYANEERREYRLRLSPDASPQQLVSTF 1533  
: | | | | | | | : | | : : : | | : | : :  
Db 1364 FSLIVPPFGKYPSLELQPMWYNEQYT-----FVSNDAP-----DTGTLELLNAL 1408

Qy 1534 RLPSGVGATCVLKS PANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPP 1593  
| | | : : | | :  
Db 1409 TKDPGFGTRCM-----EGNPI----- 1424

Qy 1594 PSPAPSDSPASPDDELQAWNVS LPTTAGPEMWT SAPSLPRLVREPVR-----C 1641  
| | | | | | | : | : : : : : |  
Db 1425 -----PD-----TPCQAGEEWT TAP-VPQTIMDLFQNGNWTM QNPSPAC 1463

Qy 1642 TCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILTDTGHNVS EYLLFTSDRF----- 1690  
| | : : | | | | | : | | | | | : | : | : | : | :  
Db 1464 QCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKN 1523

Qy 1691 ----RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK----- 1721  
| | | : | | | : | : : |  
Db 1524 KIWVNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKHKLKAKDSSADRFLNSLGREMTG 1583

Qy 1722 IAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSAS 1781  
: | : | : | | | : : : | : | | | | | | : | : | | | | | : | |  
Db 1584 LDTRNNVKVWFENKGWHAISSFLNVINAILRANLQKGE-NPSHYGITAFNHPLNLTKQQ 1642

Qy 1782 LS-LDYLLQGTDVVIAIFIIVAMSFVPASFVVFLVAEKSTKAKHLQFVSGCNP IYWLAN 1840  
| | : : | : : | : | | | | | | | : | : | | | | : | : | | : |  
Db 1643 LSEVALMTTSVDVLVSICVIFAMSFVPASFVVFLIQUERVSKAKHLQFISGVKPV IYWLSN 1702

Qy 1841 YVWDMNLNYLVPATCCVIIILFVFDLPAYTSPTNFP AVLSLFLLYGWSITPIMYPASFVFEV 1900  
: | | | | | : | | | : | | : | | | | | : | | | | | : | :  
Db 1703 FVWDMCNVVPATLVIIIFICFQQKSYVSSTNLPVLALLLLLYGWSITPLMPASFVFKI 1762

Qy 1901 PSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLME 1960  
| : | | | : | | | : | | : | | | : | | : | | | : | | : :  
Db 1763 PSTAYVVLTSVNLFIGINGSVATFVLELFT-DNKLNNINDILKSVFLIFPHFCLGRGLID 1821

Qy 1961 MAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQR 2020  
| : : : : | : : : | | | : | | | | | | | : | : : | | | :  
Db 1822 MVKNQAMADALERFGE-NRFVSPLSWDLVGRNLFAMAVEGVVFFLITVLIQYRFFIRPRP 1880

Qy 2021 MPVSTKPVED-DVDVASERQVRVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVR 2079  
: | : | | | | | | | : | : : | : | : : | | : | : :  
Db 1881 VNAKLSPLNDEDEDVRRERQRILDGGGQNDILEIKELTKIYRRK---RKPAVDRICVGIP 1937

Qy 2080 PGECFGLLG VNGAGKTSTFKMLTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDAL 2139  
| | | | | | | : | | | | : | : | : | : | : | : | : | : | :  
Db 1938 PGECFGLLG VNGAGKSSTFKMLTGDTTVTRGDAFLNKN SILSNIHEVHQNMGYCPQFDAI 1997

Qy 2140 FDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAI 2199  
: | | | : : | : : | : : | : | : | : | : | : | : | : | :  
Db 1998 TELLTGREHVEFFALLRGVPEKEVGKVGWAIKRLGLVKYGEKYAGNYS GGNKRKLSTAM 2057

Qy 2200 ALIGYPAFIFLDEPTTGMDPKARRFLWNLILD LIKTGRSVVLTSHSMEECEALCTR LAIM 2259  
| | | | : | | | | | | | | | : : | | | | | | | : | |  
Db 2058 ALIGGPPVVF LDEPTTGMDPKARRFLWNCALSVVKEGRSVVLTSHSMEECEALCTRMAIM 2117

Qy 2260 VNGRLRCLGSIQHLKNRFGDGYMITVR-TKSSQSVKDVVRFFNRFPEAMLKERHHTKVQ 2318  
 |||| |||||:||||||| | || | : :| | || || :|||:| :|  
 Db 2118 VNGRFRCLGSVQHLKNRFGDGYTIVVRIAGSNPDLKPVQDFFGLAFFPGSVLKEKHRNMLQ 2177

Qy 2319 YQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFNFAKKQSDN 2368  
 ||| | |||:| | : | | ||||| ||||| ||||| |||:  
 Db 2178 YQLPSSLSSLARIFSILSQSKRLHIEDYSVSQTTLQVFNFAKDQSDD 2227

RESULT 11

AAB31363

ID AAB31363 standard; protein; 2261 AA.

XX

AC AAB31363;

XX

DT 20-APR-2001 (first entry)

XX

DE Amino acid sequence of ABC1 polypeptide from Tangier disease patient.

XX

KW Human; adenosine triphosphate binding cassette protein 1; ABC1;

KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;

KW chromosome 9q22-9q31; heart disease; hypercholesterolemia;

KW atherosclerosis; cholesterol transport.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Misc-difference 587

FT /note= "this is changed from Arg to Trp in Tangier  
 FT disease"

XX

PN WO200078972-A2.

XX

PD 28-DEC-2000.

XX

PF 16-JUN-2000; 2000WO-US016765.

XX

PR 18-JUN-1999; 99US-0140264P.

PR 14-SEP-1999; 99US-0153872P.

PR 19-NOV-1999; 99US-0166573P.

XX

PA (CVTH-) CV THERAPEUTICS INC.

XX

PI Lawn RM, Wade D, Garvin M;

XX

DR WPI; 2001-137812/14.

XX

PT Adenosine triphosphate (ATP) binding cassette (ABC) polynucleotide,

PT useful for the development of agents for the treatment of heart disease

PT and other disorders associated with hypercholesterolemia and

PT atherosclerosis.

XX

PS Disclosure; Page 176-191; 215pp; English.

XX

CC The present sequence represents a human adenosine triphosphate (ATP)

CC binding cassette protein (ABC) 1 polypeptide, and is isolated from a

CC Tangier disease patient. ABC1 resides in cell membranes and utilises ATP

CC hydrolysis to transport a wide variety of substrates across the plasma  
 CC membrane. ABC1 is a pivotal protein in the apolipoprotein-mediated  
 CC mobilisation of intracellular cholesterol stores. ABC1 is defective in  
 CC Tangier disease, a genetic disorder characterised by abnormal HDL-  
 CC cholesterol metabolism. The ABC1 gene is localised to chromosome 9q22-  
 CC 9q31. The ABC1 genes and proteins are useful for developing  
 CC pharmaceutical agents for the treatment of heart disease and other  
 CC disorders associated with hypercholesterolemia and atherosclerosis. The  
 CC genes are useful for developing screening assays to screen for compounds  
 CC that regulate the expression of genes associated with cholesterol  
 CC transport. The genes and proteins are also useful for are also useful as  
 CC diagnostic indicators of cardiovascular disease and other disorders  
 CC associated with hypercholesterolemia

XX

SQ Sequence 2261 AA;

Query Match 33.5%; Score 4244.5; DB 4; Length 2261;  
 Best Local Similarity 39.8%; Pred. No. 1.2e-307;  
 Matches 999; Conservative 348; Mismatches 728; Indels 435; Gaps 61;

Qy	6	QLQLLWKNVTLKRRSPWVLA FEI FI PLV LFFILLGLRQKKPTISVKEVPFYTAAPL TSA	65
		:           :   :       :     :     :     :	
Db	6	QLRLLWKNLTFRRRQTCQLLLEVAVWPLFIFILILISVRLSYPPYEQHECHFPNKA-MPSA	64
Qy	66	GILPVMQSLCPDGQRDEFGFL-----QYANSTVTQLLERLDRVVEEGNLFDPARP	115
		:   : :   : :     :   :   :   :   :	
Db	65	GTLPWVQGIICNANNPCFRYPTPGEAPGVVGNFNKSIVARLFSDARRLL----LYSQKDT	120
Qy	116	SLGSELEALR--QHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSLP	173
		: :       : :   : :   :	
Db	121	SMKDMRKVLRTLQQIKKSSSNLKLQDFLVDNETFSG-----FLYHNLSLP	165
Qy	174	NSTAQALLAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPA	233
		:       :   :                 :     :	
Db	166	KSTVDKMLRADV----ILHKVFLQGYQLHLTS-LCNGS-----KSEEMI----	204
Qy	234	LLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARFSGLSAELRNQLDV	293
		: : :   : :   :	
Db	205	---QL----GDQEVSELGCLPKEKLAAAE-----RVLRSNMDI	235
Qy	294	AK-VSQQLGLDAPNGSDSSPQAPPPRRLQALLGDLLD-----AQKVLQDQDVLS	341
		: :   :     :   :   :	
Db	236	LKPILRTLNSTSPFPKELAE--TKTLLHSLGTLAQELFSMRSWSDMRQEVMTNVS	293
Qy	342	ALALLLPQGACTGRTPGPPASGAGGAAN-----GTGAGAVMGNATAEEGAPSAAALATP	396
		: :   :       : :   :         :	
Db	294	SSSSTQIYQAVSRIVCGHPEGGLKIKSLNWDYEDNNYKALFGGNGTEEDAETFYDNSTTP	353
Qy	397	---DTLQGQCSAFVQ--LWAGLQPILCGNRTIEPEALRRGNMSSLGFTSKEQRNLGLLV	451
		: :   : :       :	
Db	354	YCNDLMKNLESSPLSRIIWKALKPLLVG-----	381
Qy	452	HLMTSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVWLNISAEIRS FLEQGR LQQ	511
		: : :   :   :   :   :   :   :	
Db	382	-----KILYTPDTPATRQVMAEVNKTFOELAVFHDLEGMWEELSPKIWTFMENSQEMD	434
Qy	512	HLRWL-----QQYVAELRLHPE---ALNLSLDELPPALRQDNFS	547

Db	435	LVRMLLDSRDNDHFWEQQLDGLDWTAQDIVAFLAKHPEDVQSSNGSVYTWREAFNETN--	492
Qy	548	LPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPEESIVNYTLNQAYQDNVTVFAS	607
Db	493	-----QAIRTIS-----RMECEVNLNKLEPIATEVWLINKSME--LLDERKFWAG	535
Qy	608	VIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTG---GRFYFLYGFVW	662
Db	536	IVFTGITPGSIELPHHVKYKIRMDIDNVERTNLIKDGWDPGPRADPFEDMWYVWGGFAY	595
Qy	663	IQDMMERAIDTFVGHVVEPGSYVQMFYPYCYTRDDFLFVIEHMMPLCMVISWVYSVAM	722
Db	596	LQDVVEQAIIRVLTGTE-KKTGVYMQQMPYPCYVDDIFLRVMSRSMPLFMTLAWIYSVAV	654
Qy	723	TIQHIVAEKEHRLKEVMKTMGLNNAVHWVAFITGVQLSISVTALTALIKYQVLMHSH	782
Db	655	IIKGIVYEKEARLKETMRIMGLDNSILWFSWFISSLIPLLVSAGLLVILKLGNNLPYS	714
Qy	783	VVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHDK	842
Db	715	PSVVFVFLSVFAVVTILQCFLISTLFSRANLAAACGGIIYFTLYLPYVLC-----VAWQD	769
Qy	843	ITAFE-KCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTMLM	901
Db	770	YVGFTLKIFASLLSPVAFGFGCEYFALFEEQIGVQWDNLFESPVEEDGFNLTTISIMML	829
Qy	902	VDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSVM	961
Db	830	FDTFLYGMTWYIEAVFPGQYGIIPRPWYFPCTKSYWFG-----ESDEKSHPGSNQKRMS--	884
Qy	962	EEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTKVYKDDKKLALNKLNLNLYENQVV	1021
Db	885	---EIC-----MEEEPHTLKLGVSIQNLVKVYRDGMKVAVDGLALNFYEGQIT	929
Qy	1022	SFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLT	1081
Db	930	SFLGHNGAGKTTTMSILTGLFPPTSGTAYILGKDIRSEMSTIRQNLGVCQPQHNVLFDMLT	989
Qy	1082	VEEHLWFYSRLKSMAQEEIRREMDKMIEDLEL-SNKRHSLVQTLSGGMKRKLSVAIAFVG	1140
Db	990	VEEHIWFYARLKLSEKHVKAEMEOMALDVGLPSSKLKSKTSQLSGGMQRKLSVALAFVG	1049
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Db	1050	GSKVVILDEPTAGVDPYSRRGIWELLKRYRQRTIILSTHHMDEADVLGDRIAIISHGKL	1109
Qy	1201	KCCGSPFLKGTYG DG YRLTLVKRPAEPG-----GPQEPGLAS	1238
Db	1110	CCVGSSLFLKNQLGTGYLLTLVKKDVESSLSSCRNSSSTVSYLKKEDSVSQSSSDAGLGS	1169
Qy	1239	SPPGRAPLSSCSELQVSQFIRKHVASCLLVSDTSTELSYILPSEAAKKGA FERLFQHLER	1298
Db	1170	DHESDTLTIDVS--AISNLIRKHVSEARLVEDIGHELT YVLPYEA AKEGAFVELFHEIDD	1227
Qy	1299	SLDALHLSSFGLMDTTLEEVLKVSEEDQSLENSEADV KESRKDVLPGAEGPASGE GHAG	1358

Db	1228	RLSDLGISSYGISETTLEEIFLKVAEE-----SGVDA-ETSDGTLP-----	1267
Qy	1359	NLARCSELTQSQASLQSASSVGSARGDEGAGYTDVYGDYRPLF-DNPQDPD--NVSLQEV	1415
		:   :        :    : :: :	
Db	1268	-----ARRNRRA-FGDKQSCLRPFTEDDAADPNDSIDPESR	1303
Qy	1416	EAEALSRV-GQGSRKLDGGWLKVRQFHGLLVCRFHCARRNSKALFSQIILLPAFFVCVAMT	1474
		:    :   :    : :     :             :     :    :       :   :	
Db	1304	ETDLLSGMDGKGSYQVKGWKLTOQQFVALLWKRLLIARRSRKGFFAQIVLPAVFVCIALV	1363
Qy	1475	VALSVPEIGDLPLVLSPSQYH-NYTQPRGNFI PYANEERREYRLRLSPDASPQQIVSTF	1533
		:                  :    : : :     :    :	
Db	1364	FSLIVPPFGKYPSLELQPMWYNEQYT-----FVSNDAP-----DTGTLELLNAL	1408
Qy	1534	RLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPP	1593
		: :     :	
Db	1409	TKDPGFGTRCM-----EGNPI-----	1424
Qy	1594	PSPAPSDSPASPDEDLQAWNVS LPPTAGPEMWTSA PSLPRLVREPVR-----C	1641
		:   : : : :	
Db	1425	-----PD-----TPCQAGEEWTAP-VPQTIMDLFQNGNWTM QNPSPAC	1463
Qy	1642	TCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILTDITGHNVSEYLLFTSDRF-----	1690
		:          :        :      :   :   :   :	
Db	1464	QCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLKTYVQIIAKSLKN	1523
Qy	1691	-----RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK-----	1721
		:       :   :   :	
Db	1524	KIWNNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTG	1583
Qy	1722	IAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSAS	1781
		:   :   :      :   : :    :                :   :            :   :	
Db	1584	LDTRNNVKVWFNNKGWHAISSFLNVINNAAILRANLQKE-NPSHYGITAFNHPLNLTKQQ	1642
Qy	1782	LS-LDYLLQGTDVVIAIFIIVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLAN	1840
		: :    : :   :                 :   : :         :     :     :   :	
Db	1643	LSEVALMTTSVDVLVSICVIFAMSFVPASFVFLIQERVSKAKHLQFISGVKPVIIYWLSN	1702
Qy	1841	YVWDMNLNYLVPATCCVIIIFVFDLPAYTSPTNFPAVLSLFLLYGWSITPIMYPASFWEV	1900
		:          :      :      :         :             :      : :	
Db	1703	FVWDMCNYVVPATLVIIIFICFQKQKSYVSSTNLPVLALLLLLYGWSITPLMYPASFVFKI	1762
Qy	1901	PSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVVNSYLKSCFLIFPNYNLGHGLME	1960
		:        :        :      :         :     :         :     : :	
Db	1763	PSTAYVVLTSVNLFIGINGSVATFVLELFT-DNKLNNINDILKSVFLIFPHFCLGRGLID	1821
Qy	1961	MAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQR	2020
		: : : :   : : :       :                 :   : :         :	
Db	1822	MVKNQAMADALERFGE-NRFVSPLSWDLVGRNLFAMAVEGVVFFLITVLIQYRFFIRPRP	1880
Qy	2021	MPVSTKPVED-DVDVASERQVRVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVR	2079
		:   :             :     : : :   :   : :       :   :   :	
Db	1881	VNAKLSPLNDEDEDVRRERQRILDGGGQNDILEIKELTKIYRRK---RKPAVDRICVGIP	1937
Qy	2080	PGECFGLLVNGAGKTSTFKMLTGDESTTGGEAFVNGHSLVKELLQVQQSLGYCPQCDAL	2139
		:           :     :   :   :   :   :           :	
Db	1938	PGECFGLLVNGAGKSSTFKMLTGDTTVTRGDAFLNKNSILSNIHEVHQNMGYCPQFDAI	1997

Qy 2140 FDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAI 2199  
: || |||:: : |||: |: :| :||: || | || :| || ||||| |||||:  
Db 1998 TELLTGREHVEFFALLRGVPEKEVGKVGEWAIRKGLVKYGEKYAGNYSGGNKRKLSTAM 2057

Qy 2200 ALIGYPAFIFLDEPTTGMDPKARRFLWNLIILDLIKTGRSVVLTSHSMEECEALCTRLAIM 2259  
|||| | :||||| ||||| || | :| ||||| ||||| |||||: |||  
Db 2058 ALIGGPPVFLDEPTTGMDPKARRFLWNCALSVVKEGRSVVLTSHSMEECEALCTRMALIM 2117

Qy 2260 VNGLRLCLGSIQHLKNRFGDGYMITVR-TKSSQSVKDVVRFFNRFPEAMLKERHHTKVQ 2318  
|||| |||||: ||||| || || |: :| | || || :|||: |  
Db 2118 VNGRFRCGLGSVQHLKNRFGDGYTIVVRIAGSNPDLKPVQDFFGLAFPGSVLKEKHRNMLQ 2177

Qy 2319 YQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFVNFAKQSDN 2368  
||| | |||: || : | | ||||| ||||| ||||| |||:  
Db 2178 YQLPSSLSSLARIFSILSQSKRLHIEDYSVSQTTLDQVFVNFAKDQSDN 2227

# RESULT 12

AAB31367

ID AAB31367 standard; protein; 2261 AA.

XX

AC AAB31367;

XX

DT 20-APR-2001 (first entry)

XX

DE Amino acid sequence of ABC1 polypeptide from Tangier disease patient.

XX

KW Human; adenosine triphosphate binding cassette protein 1; ABC1;

KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;

KW chromosome 9q22-9q31; heart disease; hypercholesterolemia;

KW atherosclerosis; cholesterol transport.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Misc-difference 587

FT /note= "this is changed from Arg to Trp in Tangier

FT disease"

XX

PN WO200078971-A2.

XX

PD 28-DEC-2000.

XX

PF 16-JUN-2000; 2000WO-US016591.

XX

PR 18-JUN-1999; 99US-0140264P.

PR 14-SEP-1999; 99US-0153872P.

PR 19-NOV-1999; 99US-0166573P.

XX

PA (CVTH-) CV THERAPEUTICS INC.

PA (UNIW ) UNIV WASHINGTON.

XX

PI Lawn RM, Wade D, Oram JF, Garvin M;

XX

DR WPI; 2001-137811/14.

DR N-PSDB; AAF24708.





Db 294 SSSSTQIYQAVSRIVCGHPEGGLKIKSLNWEYEDNNYKALFGGNGTEEDAETFYDNSTTP 353  
 Qy 397 ---DTLQGQCSAFVQ--LWAGLQPILCGNNRTIEPEALRRGNMSSLGFTSKEQRNLGLLV 451  
 | :: | : : : | | : | : |  
 Db 354 YCNDLMKNLESSPLSRIIWKALKPLLVG----- 381  
 Qy 452 HLMTSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVWLNISAEIRSFLQGRLLQ 511  
 | | | | | : | : : | : | : : : | : | : : | : | :  
 Db 382 -----KILYTPDTPATRQVMAEVNKTQELAVFHDLEGMWEELSPKIWTFMENSQEMD 434  
 Qy 512 HLRWL-----QQYVAELRLHPE---ALNLSLDELPPALRQDNFS 547  
 : | | | | | | | | | | : | : | : | : | : | : | : |  
 Db 435 LVRMLLDSRDNDHFWEQQLDGLDWTQAQDIVAFLAKHPEDVQSSNGSVYTWREAFNETN-- 492  
 Qy 548 LPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQDNVTVFAS 607  
 | : : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 493 -----QAIRTIS-----RFMECVNLNKLEPIATEVWLINKSME--LLDERKFWAG 535  
 Qy 608 VIFQTRKDGS--LPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTG---GRFYFLYGFVW 662  
 : : | | | | | | | | : | : | : | : | | | : | | :  
 Db 536 IVFTGITPGSIELPHHVYKIRMDIDNVERTNKIKDGYWDPGPRADPFEDMWWYVWGGFAY 595  
 Qy 663 IQDMMERAIIDTFVGHVVEPGSYVQMFYPCYTRDDFLEFVIEHMMPLCMVISWVYSVAM 722  
 : | : : | : | | : : | : | | | | | | | : | : | : | : | : | : | :  
 Db 596 LQDVVEQAIIRVLTGTE-KKTGVYMQQMPYPCYVDDIFLRVMSRSMPLFMTLAWIYSVAV 654  
 Qy 723 TIQHIVAKEHRLKEVMKTMGLNNAVHWVAWFITGFVQLSISVTALTAILKYGOVLMHSH 782  
 | : | | | | | | | | : | : | : | : | : | : | : | : | : | : | :  
 Db 655 IIKGIVYEKEARLKETMRIMGLDNSILWFSWFISSLIPLLVSAAGLLVVILKLGNNLLPYSD 714  
 Qy 783 VWIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHDK 842  
 : : : : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 715 PSVVVFVFLSVFAVVITLQCFLISTLFSRANLAAACGGIIYFTLYLPYVLC-----VAWQD 769  
 Qy 843 ITAFE-KCIASLMSTTAFGLGSKYFALYEYVAGVGIQWHTFSQSPVEGDDFNLLLAVTMLM 901  
 | | | : | | | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 770 YVGFTLKI FASLLSPVAFGFGCEYFALFEEQIGVQWDNLFESPVEEDGFNLTTSSISMM 829  
 Qy 902 VDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSVM 961  
 | : | : : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 830 FDTFLYGVMTWYIEAVFPGQYGI PRPWYFPCTKSYWFGE---ESDEKSHPGSNQKRMS-- 884  
 Qy 962 EEDQACAMESRRFEETRGMEEPTHLPVVCVDKLTKVYKDDKKLALNKLNLNLYENQVV 1021  
 : | | | | | | | | : | : | : | : | : | : | : | : | : | : | :  
 Db 885 ---EIC-----MEEPTHLLKLVSIQNLVKVYRDGMKVAVDGLALNFYEGQIT 929  
 Qy 1022 SFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLT 1081  
 | | | | | | | | | | | | | : | | | : | | | : | : | : | : | : | : | : | :  
 Db 930 SFLGHNGAGKTTTMSILTGLFPPTSGTAYILGKDIRSEMSTIRQNLGVCPOHNVLFDM 989  
 Qy 1082 VEEHLWFYSRLKSMAQEEIRREMDKMIEDLEL-SNKRHSLVQTLSSGGMKRKLSVAIAFVG 1140  
 | : | : | : | : | : : : : | : : | : | : | : | : | : | : | : | :  
 Db 990 VEEHIWFYARLKGLSEKHVKAEMEOMALDVGLPSSKLKSKTSQSLSSGGMQRKLSVALAFVG 1049  
 Qy 1141 GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL 1200  
 | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 1050 GSKVVILDEPTAGVDPYSRRGIWELLKYRQGRITILSTHHMDEADVLDRIAIISHGKL 1109

Qy 1201 KCCGSPLFLKGTYG DG YRLTLVKRPAEPG-----GPQEPGLAS 1238  
 | | | | | | | | : | : | |  
 Db 1110 CCVGSSSLFLKNQLGTGYLTLVKKDVESLSSCRNSSSTVSYLKKEDSVSQSSSDAGLGS 1169

Qy 1239 SPPGRAPLSSCSELQVSQFIRKHAVASCLLVSDTSTELSYILPSEAAKGA FERLFQHLER 1298  
 | : | | | | : | | | | : | | | | : | | : :  
 Db 1170 DHESDTLTIDVS--AISNLIRKXHVSEARLVEDIGHELTYVLPYEAKEGAFVELFHEIDD 1227

Qy 1299 SLDALHLSSFGLMDTTLEEVFLKVSEEDQSLENSEADV KESRKDVLPGAEGPASGEGHAG 1358  
 | | : | | : : | | | | : | | | | : | | : | |  
 Db 1228 RLSDLGISSYGISSETTLEEIFLKVAEE-----SGVDA-ETSDGTLP----- 1267

Qy 1359 NLARCSELTQSQASLQSASSVGSARGDEGAGYTDVYGDYRPLF-DNPQDPD--NVSLQEV 1415  
 | | : | : | | | : | | : : :  
 Db 1268 -----ARRNRR-FGDKQSCLRPFTEDDAADPNDSIDIPESR 1303

Qy 1416 EAEALSRV-GQGSRKLDGGWLKVRQFHGLLVKRFHCARRNSKALFSQILLPAFFVVCVAMT 1474  
 | : | | : | : | : | | : | | | | : | | : | | : | :  
 Db 1304 ETDLLSGMDGKGSYQVKGWKLTTQQQFVALLWKRLLIARRSRKGFFAQIVLPAVFVCIALV 1363

Qy 1475 VALSVPEIGDLPPLVLSPSQYH-NYTQPRGNFIPIYANEERREYRLRLSPDASPQQLVSTF 1533  
 : | | | | | | : | | : : : | | : | :  
 Db 1364 FSLIVPPFGKYPSPLELQPMWYNEQYT-----FVSNDAP E-----DTGTLELLNAL 1408

Qy 1534 RLPSGVGATCVLKSPANGSLGPTLNLSGESSRLLAARFFDSMCLESFTQGLPLSNFVPPP 1593  
 | | | : | | : | | : | | : | | :  
 Db 1409 TKDPGFGTRCM-----EGNPI----- 1424

Qy 1594 PSPAPSDSPASPDLDQAWNVS LPPTAGPEMWT SAPSLPRLVREPVR-----C 1641  
 | | | | | | | | : | | : : : : : |  
 Db 1425 -----PD-----TPCQAGEEEWTTAP-VPQTIMDLFQNGNWTM QNPSPAC 1463

Qy 1642 TC SAQGTGFS---CPSSVGG-HPPQMRVVTGDILTDTIGHNVSEYLLFTSDRF----- 1690  
 | | : | | | | | | | : | | | | : | | : | | : | :  
 Db 1464 QCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNIDYLVKTYVQIIAKSLKN 1523

Qy 1691 ----RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK----- 1721  
 | | | : | | | : | | : : |  
 Db 1524 KIWVNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKHKLAKDSSADRFLNSLGRFMTG 1583

Qy 1722 IAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHMPMNTSAS 1781  
 : | : | : | | | : | | : | | | | | : | | : | | : | |  
 Db 1584 LDTRNNVKVWFENKGWHAISSFLNVINNA ILRANLQKGE-NPSHYGITAFNHPLNLTKQQ 1642

Qy 1782 LS-LDYLLQGT DVVIAIFIIVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLAN 1840  
 | | : : | | : | | | | | | | : | : | | | | : | | : | | : |  
 Db 1643 LSEVALMTTSVDVLVSICVIFAMSFVPASFVFLIQERVSKAKHLQFISGVKPVIIYWLSN 1702

Qy 1841 YVWDMNLNYLVPATCCVILFVFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWFEV 1900  
 : | | | | | : | | | : | | | : | | | | | : | | : | : :  
 Db 1703 FVWDMCNYVVPATLVIIIFICFQQKSYVSSNLPVLALLLLLYGWSITPLMYPASFVFKI 1762

Qy 1901 PSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLME 1960  
 | | : | | | : | | | : | | | : | | | | : | | : | : :  
 Db 1763 PSTAYVVLTSVNLFIGINGSVATFVLELFT-DNKLNNINDILKSVFLIFPHFCLGRGLID 1821

Qy 1961 MAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQR 2020  
 | |: : : : |: : : || ||:| | | ||||| ||:|: : || | ||:  
 Db 1822 MVKNQAMADALERFGE-NRFVSPLSWDLVGRNLFAMAVEGVVFFLITVLIQYRFFIRPRP 1880

Qy 2021 MPVSTKPVED-DVDVASERQVRVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRCLGVR 2079  
 : | : | | | | |||:| | ||::|: |||:|: : | |||:|:|:  
 Db 1881 VNAKLSPLNDEDEDVRRERQRILDGGGQNDILEIKELTKIYRRK---RKPAVDRICVGIP 1937

Qy 2080 PGECFGLLG VNGAGKTSTFKMLTGDESTTGGEAFVNGHSLVKELLQVQQSLGYCPQCDAL 2139  
 ||||| |||||:||||| : | |:|:| :|:| : :| |:| |||| ||:  
 Db 1938 PGECFGLLG VNGAGKSSTFKMLTGDTTVTRGDAFLNKNLSILSNIHEVHQNMGYCPQFDAI 1997

Qy 2140 FDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAI 2199  
 : || |||: : |||: |: :| :|: || | || :| || ||||| |||||:  
 Db 1998 TELLTGREHVEFFALLRGVPEKEVGKVGWAIKRLGLVKYGEKYAGNYSGGNKRKLSTAM 2057

Qy 2200 ALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIM 2259  
 |||| | :||||| ||||| | :| ||||| ||||| |||||: |||||:  
 Db 2058 ALIGGPPVFLDEPTTGMDPKARRFLWNCALSVVKEGRSVVLTSHSMEECEALCTRMAIM 2117

Qy 2260 VNGRLRCLGSIQHLKNRFGDGYMITVR-TKSSQSVKDVVRFFNRFPEAMKERHHTKVQ 2318  
 |||| ||||:||||| | || |: :| | || || :|||:| :|  
 Db 2118 VNGRFRCLG SVQHLKNRFGDGYTIVVRIAGSNPDLKPVQDFFGLAFPGSVLKEKHRNMLQ 2177

Qy 2319 YQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFNFAKKQSDN 2368  
 ||| | |||:|: | | ||||| ||||| ||||| |||:  
 Db 2178 YQLPSSLSSLARIFSILSQSKRLHIEDYSVSQTTLQVFNFAKDQSD 2227

RESULT 13

AAB38109

ID AAB38109 standard; protein; 2261 AA.

XX

AC AAB38109;

XX

DT 29-JAN-2001 (first entry)

XX

DE Human ABC1 cholesterol transporter mutant, R219K.

XX

KW Human ABC1 cholesterol transporter; chromosome 9q31;

KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;

KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;

KW cardiovascular disease; coronary artery disease; coronary restenosis;

KW cerebrovascular disease; peripheral vascular disease;

KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;

KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;

KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;

KW mutein.

XX

OS Homo sapiens.

XX

PN WO200055318-A2.

XX

PD 21-SEP-2000.

XX

PF 15-MAR-2000; 2000WO-IB000532.

XX

PR 15-MAR-1999; 99US-0124702P.  
PR 08-JUN-1999; 99US-0138048P,  
PR 17-JUN-1999; 99US-0139600P.  
PR 01-SEP-1999; 99US-0151977P.

XX

PA (UYBR-) UNIV BRITISH COLUMBIA.  
PA (XENO-) XENON BIORESEARCH INC.

XX

PI Hayden MR, Wilson AR, Pimstone SN;

XX

DR WPI; 2000-587528/55.

XX

PT New ABC1 polypeptide is useful for treating diseases associated with ABC1  
PT biological activity, e.g. Alzheimer's disease, Huntington's disease and  
PT cancer.

XX

PS Example; Page; 229pp; English.

XX

CC The invention relates to the human ABC1 cholesterol transporter protein  
CC (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is  
CC a member of the ATP-binding cassette (ABC transporter) superfamily of  
CC proteins, and plays a crucial role in cholesterol transport, particularly  
CC intracellular cholesterol trafficking in monocytes and fibroblasts, being  
CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is  
CC located on chromosome 9q31, and mutations in this gene are associated  
CC with two genetic HDL (high density lipoprotein) deficiency disorders,  
CC Tangier disease (TD) and familial HDL deficiency (FHA). These diseases  
CC are distinguishable in that TD is an autosomal recessive disorder, while  
CC FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good  
CC cholesterol") in the blood correlate with a high risk of cardiovascular  
CC disease, particularly coronary artery disease, but also cerebrovascular  
CC disease, coronary restenosis, and peripheral vascular disease.  
CC Conversely, a high level of HDL has protective effects against  
CC cardiovascular disease. The invention provides genetic constructs and  
CC transgenic cells and non-human animals comprising human ABC1 nucleic  
CC acids, and methods of gene therapy for the treatment or prevention of  
CC cardiovascular disease comprising the administration of an expression  
CC vector encoding ABC1 or an active fragment thereof. The invention also  
CC encompasses compounds which mimic ABC1 activity, compounds which  
CC stimulate ABC1 expression and methods of screening for such compounds. It  
CC further relates to methods for determining whether a patient has an  
CC increased risk for cardiovascular disease due to polymorphisms in the  
CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat or  
CC prevent cardiovascular disease, especially coronary artery disease,  
CC cerebrovascular disease, coronary restenosis or peripheral vascular  
CC disease. They may also be used in the treatment of diseases associated  
CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick  
CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.  
CC The invention specifically excludes proteins with the exact amino acid  
CC sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic  
CC acid with the exact sequence as GenBank Accession No: AJ012376.1. The  
CC present sequence represents a mutant human ABC1 cholesterol transporter  
CC associated with an altered cholesterol level and therefore an altered  
CC risk of cardiovascular disease. Note: The present sequence is not shown  
CC in the specification, but is derived from the native human ABC1 shown on  
CC pages 152-157

XX

SQ Sequence 2261 AA;

Query Match 33.5%; Score 4241.5; DB 3; Length 2261;  
Best Local Similarity 39.8%; Pred. No. 2.1e-307;  
Matches 1000; Conservative 346; Mismatches 729; Indels 435; Gaps 61;

```
Qy      6 QLQLLLWKNVTLKRRSPWVLAFEIFIPLVLFILLGLRQKKPTISVKEVPFYTAAPL TSA 65
      ||:|||||:| :|| | | : || :| ||: :| | | | : ||
Db      6 QLRLLLWKNLTFRRRQTCQLLLEVAWPLFIFLILISVRLSYPPYEQHECHFNPKA-MPSA 64

Qy     66 GILPVMQSLCPDQORDEFQFL-----QYANSTVTQLLERLDRVVEEGNLFDPARP 115
      | || :| : : | : : | | :| | :| | :| :|
Db     65 GTLPWVQGIICNANNPCFRYPTPGEAPGVVGNFNKSIVARLFSDARRLL----LYSQKDT 120

Qy    116 SLGSELEALR--QHLEALSAGPGTSGSHLDRSTVSSFSLDVARNPQELWRFLTQNLSLP 173
      |: : || | : : |: : | | | | | | | | | | | | | |
Db    121 SMKDMRKVLRTLQQIKKSSSNLKLQDFLVDNETFSG-----FLYHNLSLP 165

Qy    174 NSTAQALLAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPA 233
      || :| | | :| :| | | | | | | | | | | | | |
Db    166 KSTVDKMLRADV----ILHKVFLQGYQLHLTS-LCNGS-----KSEEMI---- 204

Qy    234 LLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARFSGLSAELRNQLDV 293
      || | | : : :| : : | : | | | | | | | | | |
Db    205 ---QL----GDQEVSELCLPKEKLAAAE-----RVLRSNMDI 235

Qy    294 AK-VSQQLGLDAPNGSDSSPQAPPPRRLQALLGDLLD-----AQKVLQDQDVLS 341
      | : : | :| | :| :| | | | | | | | | | | | |
Db    236 LKPILRTLNSTSPFPKELAAE--TKTLLHSLGTLAQELFSMRSWSDMRQEVMTNVNS 293

Qy    342 ALALLLPQGACTGRTPGPPASGAGGAAN-----GTGAGAVMGNPATAEEGAPSAAALATP 396
      : : | : | | | : : | | | | | : | | | | | |
Db    294 SSSSTQIYQAVSRIVCGHPEGGLKIKSLNWEYEDNNYKALFGGNGTEEDAETFYDNSTTP 353

Qy    397 ---DTLQGQCSAFVQ--LWAGLQPILCGNRTIEPEALRRGNMSSLGFTSKEQRNLGLLV 451
      | : : | : : :| | :| :| | |
Db    354 YCNDLMKNLESSPLSRIIWKALKPLLVG----- 381

Qy    452 HLMTSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVWLNISAEIRS FLEQGR LQ 511
      |||| | : : : | :| : : | :| :| :| :| :| :
Db    382 -----KILYTPDTPATRQVMAEVNKTQELAVFHDLEGMWEELSPKIWTFMENSQEMD 434

Qy    512 HLRWL-----QQYVAELRLHPE---ALNLSLDELPPALRQDNFS 547
      :| | | | | | | | | | | | | | | | | | |
Db    435 LVRMLLDSRDNDHFWEQQLDGLDWTQAQDIVAFLAKHPEDVQSSNGSVYTWREAFNETN-- 492

Qy    548 LPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQDNVTVFAS 607
      | : || :|| | : : : | :| :| :| | :| :|
Db    493 -----QAIRTIS-----RFMECVNLNKLEPIATEVWLINKSME--LLDERKFWAG 535

Qy    608 VIFQTRKDGS--LPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTG---GRFYFLYGFVW 662
      : :| || || || || || : | :| :| :| || || | || :
Db    536 IVFTGITPGSIELPHHVYKIRMDIDNVERTNKKIDGYWDPGPRADPFEDMRYVWGGFAY 595

Qy    663 IQDMMERAIIDTFVGHDVVEPGSYVQMFPYPCYTRDDFLFVIEHMMPLCMVISWVYSVAM 722
      :||:|:|:| | : : | :| |||| | || : | | :| :| :| :|
Db    596 LQDVVEQAIIRVLTGTE-KKTGVYMQMPYPCYVDDIFLRVMSRSMPLFMTLAWIYSVAV 654
```

Qy 723 TIQHIVAEKEHRLKEVMKTMGLNNAVHWVAWFITGQVLSISVTALTALIKYQVLMHSH 782  
 Db 655 IIKGIVYEKEARLKETMRIMGLDNSILWFSWFISSLIPLLVSAGLLVILKLGNNLLPYSD 714

Qy 783 VVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHDK 842  
 Db 715 PSVVFVFLSVFAVVTILQCFLISTLFSRANLAAACGGIIYFTLYLPYVLC-----VAWQD 769

Qy 843 ITAFE-KCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTMLM 901  
 Db 770 YVGFTLKIFASLLSPVAFGFGCEYFALFEEQGIGVQWDNLFESPVEEDGFNLTTSVSMML 829

Qy 902 VDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSVM 961  
 Db 830 FDTFLYGVMTWYIEAVFPGQYGI PRPWYFPCTKSYWFGE---ESDEKSHPGSNQKRIS-- 884

Qy 962 EEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTQVYKDDKKLALNKLNLNLYENQVV 1021  
 Db 885 ---EIC-----MEEEPHTLKLGVSIQNLVKVYRDGMKVAVDGLALNFYEGQIT 929

Qy 1022 SFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLT 1081  
 Db 930 SFLGHNGAGKTTTMSILTGLFPPTSGTAYILGKDIRSEMSTIRQNLGVCQHNVLFDMLT 989

Qy 1082 VEEHLWFYSRLKSMAQEEIRREMDKMIEDLEL-SNKRHSIVQTLSSGGMKRKLSVAIAFVG 1140  
 Db 990 VEEHIWFYARLKLSEKHVKAEMEOMALDVGLPSSKLKSKTSQSLSGGMQRKLSVALAFVG 1049

Qy 1141 GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTIILSTHHMDEADLLGDRIAIISHGKL 1200  
 Db 1050 GSKVVILDEPTAGVDPYSRRGIWELLLKYRQGRTIILSTHHMDEADVLGDRIAIISHGKL 1109

Qy 1201 KCCGSPLFLKGTYGDRYTLVLRPAEPG-----GPQEPGLAS 1238  
 Db 1110 CCVGSSLFLKNQLGTGYTLTVKKDVESLSSCRNSSSTVSYLKKEDSVSQSSSDAGLGS 1169

Qy 1239 SPPGRAPLSSSELQVSQFIRKHVASCLLVSDTSTELSYILPSEAAKGAERLRFQHLER 1298  
 Db 1170 DHESDTLTIDVS--AISNLIRKHVSEARLVEDIGHELTYVLPYEAAKEGAFVELFHEIDD 1227

Qy 1299 SLDALHLSSFGLMDTTLEEVFLKVSEEDQSLENSEADVKESSKRDVLPGAEGPASGEGHAG 1358  
 Db 1228 RLSDLGISSYGISETTLEEIFLKVAEE-----SGVDA-ETSDGTLP----- 1267

Qy 1359 NLARCELTQSQASLQSASSVGSARGDEGAGYTDVYGDYRPLF-DNPQDPD--NVSLQEV 1415  
 Db 1268 -----ARRNRRA-FGDKQSCLRPFTEDDAADPNDSIDPESR 1303

Qy 1416 EAEALSRV-GQGSRLDGGWLKVRQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMT 1474  
 Db 1304 ETDLLSGMDGKGSYQVKGWKLTTQQQFVALLWKRLLIARRSRKGFFAQIVLPAVFCIALV 1363

Qy 1475 VALSVPEIGDLPPLVLSPSQYH-NYTQPRGNFIPIYANEERREYRLRLSPDASPQQLVSTF 1533  
 Db 1364 FSLIVPPFGKYPSLELQPWMYNEQYT-----FVSNDAP-----DTGTLELLNAL 1408

Qy	1534	RLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPP	1593
Db	1409	TKDPGFGTRCM-----EGNPI-----	1424
Qy	1594	PSPAPSDSPASPDEDLQAWNVSLEPPTAGPEMWTSAFSLPRLVREPVR-----C	1641
Db	1425	-----PD-----TPCQAGEEWTAP-VPQTIMDLFQNGNWTMQNPSPAC	1463
Qy	1642	TCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILTDTIGHNVSEYLLFTSDRF-----	1690
Db	1464	QCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKN	1523
Qy	1691	----RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK-----	1721
Db	1524	KIWNNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKHKLAKDSSADRFLNSLGRFMTG	1583
Qy	1722	IAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSAS	1781
Db	1584	LDTRNNVKVWFNNKGWHAISSFLNVINNAAILRANLQKGE-NPSHYGITAFNHPNLTKQQ	1642
Qy	1782	LS-LDYLLQGTDVVIAIFIIVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLAN	1840
Db	1643	LSEVALMTTSVDVLVSICVIFAMSFVPASFVFLIQERVSKAKHLQFISGVKPIYWLNS	1702
Qy	1841	YVWDMNLNVLVPATCCVILFVFDLPAYTSPTNFPVAVLSLFLLYGWSITPIMYPASFVFEV	1900
Db	1703	FVWDMCNVYVVPATLVIIIFICFQQKSYSVSTNLPVLALLLLLYGWSITPLMYPASFVKI	1762
Qy	1901	PSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLME	1960
Db	1763	PSTAYVVLTSVNLFIGINGSVATFVLELEFT-DNKLNNINDILKSVFLIFPHFCLGRGLID	1821
Qy	1961	MAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQR	2020
Db	1822	MVKNQAMADALERFGE-NRFVSPLSWDLVGRNLFAMAVEGVVFFLITVLIQYRFFIRPRP	1880
Qy	2021	MPVSTKPVED-DVDVASERQVRVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVR	2079
Db	1881	VNAKLSPLNDEDEDVRRERQRILDGGGQNDILEIKELTKIYRRK---RKPAVDRICVGIP	1937
Qy	2080	PGEFCGLLGUNGAGKTSTFKMLTGDESTTGGEAFVNGHSLVKELLQVQQLGYCPQCDAL	2139
Db	1938	PGEFCGLLGUNGAGKSSTFKMLTGDTTVTRGDAFLNKNSILSNIHEVHQNMGYCPQFDAI	1997
Qy	2140	FDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAI	2199
Db	1998	TELLTGREHVEFFALLRGVPEKEVGKVGWEAIRKGLLVKYGEKYAGNYSGGNKRKLSTAM	2057
Qy	2200	ALIGYPAFIFLDEPTTGMDPKARREFLWNLIILDIKTGRSVVLTSHSMEECEALCTRLAIM	2259
Db	2058	ALIGGPPVVFLDEPTTGMDPKARREFLWNCALSVVKEGRSVVLTSHSMEECEALCTRMAM	2117
Qy	2260	VNGRLRCLGSIQHLKNRFGDGYMITVR-TKSSQSVKDVVRFFNRFPEAMLERHHTKVQ	2318
Db	2118	VNGRFRCLGSVQHLKNRFGDGYTIVVRIAGSNPDLKPVQDFFGLAFPGSVLKEKHRNMLQ	2177
Qy	2319	YQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSOTTLDNVFVNFACKOSDN	2368



Db 2178 YQLPSSLSSLARIFSILSQSKRLHIEDYSVSQTTLDQVFNFAKDQSD 2227

RESULT 14

AAB38117

ID AAB38117 standard; protein; 2261 AA.

XX

AC AAB38117;

XX

DT 29-JAN-2001 (first entry)

XX

DE Human ABC1 cholesterol transporter mutant, I883M.

XX

KW Human ABC1 cholesterol transporter; chromosome 9q31;

KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;

KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;

KW cardiovascular disease; coronary artery disease; coronary restenosis;

KW cerebrovascular disease; peripheral vascular disease;

KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;

KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;

KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;

KW mutein.

XX

OS Homo sapiens.

XX

PN WO200055318-A2.

XX

PD 21-SEP-2000.

XX

PF 15-MAR-2000; 2000WO-IB000532.

XX

PR 15-MAR-1999; 99US-0124702P.

PR 08-JUN-1999; 99US-0138048P.

PR 17-JUN-1999; 99US-0139600P.

PR 01-SEP-1999; 99US-0151977P.

XX

PA (UYBR-) UNIV BRITISH COLUMBIA.

PA (XENO-) XENON BIORESEARCH INC.

XX

PI Hayden MR, Wilson AR, Pimstone SN;

XX

DR WPI; 2000-587528/55.

XX

PT New ABC1 polypeptide is useful for treating diseases associated with ABC1

PT biological activity, e.g. Alzheimer's disease, Huntington's disease and

PT cancer.

XX

PS Example; Page; 229pp; English.

XX

CC The invention relates to the human ABC1 cholesterol transporter protein

CC (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is

CC a member of the ATP-binding cassette (ABC transporter) superfamily of

CC proteins, and plays a crucial role in cholesterol transport, particularly

CC intracellular cholesterol trafficking in monocytes and fibroblasts, being

CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is

CC located on chromosome 9q31, and mutations in this gene are associated

with two genetic HDL (high density lipoprotein) deficiency disorders, Tangier disease (TD) and familial HDL deficiency (FHA). These diseases are distinguishable in that TD is an autosomal recessive disorder, while FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good cholesterol") in the blood correlate with a high risk of cardiovascular disease, particularly coronary artery disease, but also cerebrovascular disease, coronary restenosis, and peripheral vascular disease. Conversely, a high level of HDL has protective effects against cardiovascular disease. The invention provides genetic constructs and transgenic cells and non-human animals comprising human ABC1 nucleic acids, and methods of gene therapy for the treatment or prevention of cardiovascular disease comprising the administration of an expression vector encoding ABC1 or an active fragment thereof. The invention also encompasses compounds which mimic ABC1 activity, compounds which stimulate ABC1 expression and methods of screening for such compounds. It further relates to methods for determining whether a patient has an increased risk for cardiovascular disease due to polymorphisms in the ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat or prevent cardiovascular disease, especially coronary artery disease, cerebrovascular disease, coronary restenosis or peripheral vascular disease. They may also be used in the treatment of diseases associated with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer. The invention specifically excludes proteins with the exact amino acid sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic acid with the exact sequence as GenBank Accession No: AJ012376.1. The present sequence represents a mutant human ABC1 cholesterol transporter associated with an altered cholesterol level and therefore an altered risk of cardiovascular disease. Note: The present sequence is not shown in the specification, but is derived from the native human ABC1 shown on pages 152-157

SQ Sequence 2261 AA;

Query Match 33.5%; Score 4240.5; DB 3; Length 2261;  
Best Local Similarity 39.8%; Pred. No. 2.5e-307;  
Matches 1000; Conservative 345; Mismatches 730; Indels 435; Gaps 61;

Qy	6	QLQLLLWKNVTILKRSPWVLAFEIPIPLVLFFILLGLRQKKPTISVKEVPFYTAAPLTSA	65
Db	6	QLRLLLWKNLTFRRRQTCQLLLEVAWPLFIFLILISVRLSYPPYEQHECHFNPKA-MPSA	64
Qy	66	GILPVMQSLCPDQGRDEFGL-----QYANSTVTQLLERLDRVVEEGNLFDPARP	115
Db	65	GTLPVWQGIICNANNPCFRYPTPGEAPGVVGNFNKSIVARLFSDDARRLL----LYSQKDT	120
Qy	116	SLGSELEALR--QHLEALSAGPGTSGSHLDRSTVSSFSLDVARNPQELWRFLTQNLSLP	173
Db	121	SMKDMRKVLRTLQQIKKSSSNLKLQDFLVDNETFSG-----FLYHNLSLP	165
Qy	174	NSTAQALLAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPA	233
Db	166	KSTVDKMLRADV----ILHKVFLQGYQLHLTS-LCNGS-----KSEEMI----	204
Qy	234	LLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARFSGLSAELRNQLDV	293
Db	205	---OL----GDOEVSELCLGPKEKLAAAE-----RVLRSNMDI	235

Qy 294 AK-VSQQGLDAPNGSDSSPQAPPPRRLQALLGDLLD-----AQKVLQDQDVLS 341  
| : : | : | : | : | | | | : : : | |  
Db 236 LKPILRTLNSTSPFPSKELAEA--TKTLLHSLGTLAQELFSMRSWSDMRQEVMTLVNS 293

Qy 342 ALALLLPQGACTGRTPGPPASGAGGAAN-----GTGAGAVMGNPNATAEEGAPSAAALATP 396  
: : | : | | : : | : | | : | |  
Db 294 SSSSTQIYQAVSRIVCGHPEGGLKIKSLNWWYEDNNYKALFGGNGTEEDAETFYDNSTTP 353

Qy 397 ---DTLQGQCSAFVQ--LWAGLQPILCGNNRTIEPEALRRGNMSSLGFTSKEQRNLGLLV 451  
| : : | : : | : | : | : |  
Db 354 YCNDLMKNLESSPLSRIIWKALKPLLVG----- 381

Qy 452 HLMTSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVVLNISAEIRSFLEQGRLLQ 511  
| | | | | : : : | : | : : | : | : : | :  
Db 382 -----KILYTPDTPATRQVMAEVNKTQELAVFHDLEGMWEELS PKIWTFMENSQEMD 434

Qy 512 HLRWL-----QQYVAELRLHPE---ALNLSLDELPPALRQDNFS 547  
: | | | | | | | | | : | : | : |  
Db 435 LVRMLLDSRDNDHFWEQQLDGLDWTQAQDIVAFLAKHPEDVQSSNGSVYTWREAFNETN-- 492

Qy 548 LPSGMALLQQLDITDINAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQDNVTFAS 607  
| : | | : | | : : : | : : : | : |  
Db 493 -----QAIRTIS-----RFMECVNLNKLEPIATEVWLINKSME--LLDERKFWAG 535

Qy 608 VIFQTRKDG--LPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTG---GRFYFLYGFVW 662  
: : | | | | | | : : | : : | : | | | | :  
Db 536 IVFTGITPGSIELPHHVYKIRMDIDNVERTNKKIDGYWDPGPRADPFEDMRYVWGGFAY 595

Qy 663 IQDMMERAIIDTFVGHVVEPGSYVQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSVAM 722  
: | : : | : | | : : | : | | | | | | : | : | : | : | : | :  
Db 596 LQDVVEQAIIRVLTGTE-KKTGVYMQMPYPCYVDDIFLRVMSRSMPLFMTLAWIYSVAV 654

Qy 723 TIQHIVAEKEHRLKEVMKTMGLNNAVHVVAFWTFITGFVQLSISVTALTALIKYQVLMHSH 782  
| : | | | | | | : | : | : | : | : | : | : | : | : | : | :  
Db 655 IIKGIVYEKEARLKETMRIMGLDNSILWFSWFISSLIPLLVSAGLLVVLKLGNNLLPYSD 714

Qy 783 VVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHDK 842  
: : : : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 715 PSVVFVFLSVFAVVTILQCFLISTLFSRANLAAACGGIIYFTLYLPYVLC-----VAWQD 769

Qy 843 ITAFE-KCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTMLM 901  
| | | : | | | | : | : | : | : | : | : | : | : | : | : | :  
Db 770 YVGFTLKFASLLSPVAFGFGCEYFALFEEQGIGVQWDNLFESPVEEDGFNLTTSVSMML 829

Qy 902 VDAVYIGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSVM 961  
| : | : : | | | | | | | : | : | : | : | : | : | : | : | :  
Db 830 FDTFLYGVMTWYIEAVFPGQYGI PRPWYFPCTKSYWFG---ESDEKSHPGSNQKRMS-- 884

Qy 962 EEDQACAMESRRFEETRGMEEEPTHLPLVVCVDKLTKVYKDDKKLALNKLNLNLYENQVV 1021  
: | | | | | | | : | : | : | : | : | : | : | : | :  
Db 885 ---EIC-----MEEEPHTLKLGVSIQNLVKVYRDGMKVAVDGLALNFYEGQIT 929

Qy 1022 SFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLT 1081  
| | | | | | | | | | | | | : | : | : | : | : | : | : | : | :  
Db 930 SFLGHNGAGKTTTMSILTGLFPPTSGTAYILGKDIRSEMSTIRQNLGVCPQHNVLFDMILT 989

Qy	1082	VEEHLWFYSRLKSMQEEIRREMDKMIEDLEL-SNKRHSLVQTLSSGGMKRKLSVAIAFVG	1140
Db	990	VEEHWFYARLKGLSEKHVKAEMEQLMDVGLPSSSKLKSQTSQSLSSGGMQRKLSVALAFVG	1049
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Db	1050	GSKVVILDEPTAGVDPYSRRGIWELLLKYRQGRITILLSTHHMDEADVLGDRIAIISHGKL	1109
Qy	1201	KCCGSPLFLKGTGYDGYRLTLVKRPAEPG-----GPQEPGLAS	1238
Db	1110	CCVGSLSLFLKNQLGTGYLTLVKKDVESLSSCRNSSSTVSYLKKEDSVSQSSSDAGLGS	1169
Qy	1239	SPPGRAPLSSCSELQVSQFIRKHHVASCLLVSDTSTELSYILPSEAAKKGAFERLFQHLER	1298
Db	1170	DHESDTLTIDVS--AISNLIRKHVSEARLVEDIGHELTIVLPYEAKEGAFVELFHEIDD	1227
Qy	1299	SLDALHLSSFGLMDTTLEEVLKVSEEDQSLSENSEADVKESSKRDVLPGAEGPASGEGHAG	1358
Db	1228	RLSDLGISSYGISETTLEEIFLKVAEE-----SGVDA-ETSDGTLP-----	1267
Qy	1359	NLARCSELTQSQASLQSASSVGSARGDEGAGYTDVYGDYRPLF-DNPQDPD--NVSLQEV	1415
Db	1268	-----ARRNRRRA-FGDKQSCLRPFTEDDAADPNDSIDIPESR	1303
Qy	1416	EAEALSRV-GQGSRKLDGGWLKVRQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMT	1474
Db	1304	ETDLLSGMDGKGSYQVKGWKLTTQQQFVALLWKRLLIARRSRKGFFAQIVLPAVFCIALV	1363
Qy	1475	VALSVPEIGDLPLPLVLSPSQYH-NYTQPRGNFIPIYANEERREYRLRLSPDASPQQLVSTF	1533
Db	1364	FSLIVPPFGKYPSLELQPMWYNEQYT-----FVSNDAPF-----DTGTLELLNAL	1408
Qy	1534	RLPSGVGATCVLKSPANGSLGPTLNLSSGESRLAARFFDSMCLESFTQGLPLSNFVPPP	1593
Db	1409	TKDPGFGTRCM-----EGNPI-----	1424
Qy	1594	PSPAPSDSPASPDEDLQAWNVSLPPTAGPEMWTSAFSLPRLVREPVR-----C	1641
Db	1425	-----PD-----TPCQAGEEETAP-VPQTIMDLFQNGNWTMNPSPAC	1463
Qy	1642	TCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILTDTIGHNVSEYLLFTSDRF-----	1690
Db	1464	QCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKN	1523
Qy	1691	----RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK-----	1721
Db	1524	KIWNNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTG	1583
Qy	1722	IAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSAS	1781
Db	1584	LDTRNNVKVWFNNKGWHAISSFLNVINNAAILRANLQKGE-NPSHYGITAFNHPLNLTKQQ	1642
Qy	1782	LS-LDYLLQGTDVVIAIFIIVAMSFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLAN	1840
Db	1643	LSEVALMTTSVDVLVSIKVFAMSFVPASFVFLIQERVSKAKHLQFISGVKPIYWLSN	1702
Qy	1841	YVWDMNLNYLVPATCCVILFVFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFVFEV	1900

```

      :||| ||:|||| :|| | :| || | : | |||||:||||| |:
Db      1703 FVWDMCNYVVPATLVIIIFICFQQKSYVSSTNLPVLALLLLLYGWSITPLMYPASFVFKI 1762

Qy      1901 PSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLME 1960
      ||:||| | :||||| :||||:|:| | | :| ||| ||||:| ||:|
Db      1763 PSTAYVVITSVNLFIGINGSVATFVLELFT-DNKLNNINDILKSVFLIFPHFCLGRGLID 1821

Qy      1961 MAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQR 2020
      | |: : : : |: :| || ||:| | | ||||| ||:|:| | | ||:
Db      1822 MVKNQAMADALERFGE-NRFVSPSLWDLVGRNLFAMAVEGVVFFLITVLIQYRFFIRPRP 1880

Qy      2021 MPVSTKPVED-DVDVASERQVRVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRICLGVR 2079
      : | | | | | |||:| | ||:|:| |||:| : | |||:|:|
Db      1881 VNAKLSPLNDEDEDVRRERQRILDGGGQNDILEIKELTKIYRRK---RKPAVDRICVGIP 1937

Qy      2080 PGECFGLLVNGAGKTSTFKMLTGDESTTGGEAFVNGHSLVKELLQVQQSLGYCPQCDAL 2139
      ||||| |||||:||||||| : | |:|:| |:| : :| |:| |||| ||:
Db      1938 PGECFGLLVNGAGKSSTFKMLTGDTTVTRGDAFLNKNLSILSNIHEVHQNMGYCPQFDAI 1997

Qy      2140 FDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAI 2199
      : || |||: : |||: |: :| :|: || | | :| || ||||| |||:
Db      1998 TELLTGREHVEFFALLRGVPEKEVGKVGEWAIRKGLVKYGEKYAGNYSGGNKRKLSTAM 2057

Qy      2200 ALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIM 2259
      |||| | :||||| ||||| | :| ||||| |||||:|
Db      2058 ALIGGPPVFLDEPTTGMDPKARRFLWNCALSVVKEGRSVVLTSHSMEECEALCTRMAIM 2117

Qy      2260 VNGRLRCLGSIQHLKNRFGDGYMITVR-TKSSQSVKDVVRFFNRNFP EAMLERHHTKVQ 2318
      |||| ||||:||||||| | || |: :| | | || :|||:| :|
Db      2118 VNGRFRCLGSVQHLKNRFGDGYTIVVRIAGSNPDLKPVDFFGLAFPGSVLKEKHRNMLQ 2177

Qy      2319 YQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFVNFAKKQSDN 2368
      ||| | |||:| : | | ||||| ||||| |||:
Db      2178 YQLPSSLSSLARIFSILSQSKRLHIEDYSVSQTTLDQVFNFAKDQSD 2227

```

RESULT 15

AAB38114

ID AAB38114 standard; protein; 2261 AA.

XX

AC AAB38114;

XX

DT 29-JAN-2001 (first entry)

XX

DE Human ABC1 cholesterol transporter mutant, E1172D.

XX

KW Human ABC1 cholesterol transporter; chromosome 9q31;

KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;

KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;

KW cardiovascular disease; coronary artery disease; coronary restenosis;

KW cerebrovascular disease; peripheral vascular disease;

KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;

KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;

KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;

KW mutein.

XX

OS Homo sapiens.

XX  
 PN WO200055318-A2.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 15-MAR-2000; 2000WO-IB000532.  
 XX  
 PR 15-MAR-1999; 99US-0124702P.  
 PR 08-JUN-1999; 99US-0138048P.  
 PR 17-JUN-1999; 99US-0139600P.  
 PR 01-SEP-1999; 99US-0151977P.  
 XX  
 PA (UYBR-) UNIV BRITISH COLUMBIA.  
 PA (XENO-) XENON BIORESEARCH INC.  
 XX  
 PI Hayden MR, Wilson AR, Pimstone SN;  
 XX  
 DR WPI; 2000-587528/55.  
 XX  
 PT New ABC1 polypeptide is useful for treating diseases associated with ABC1  
 PT biological activity, e.g. Alzheimer's disease, Huntington's disease and  
 PT cancer.  
 XX  
 PS Example; Page; 229pp; English.  
 XX  
 CC The invention relates to the human ABC1 cholesterol transporter protein  
 CC (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is  
 CC a member of the ATP-binding cassette (ABC transporter) superfamily of  
 CC proteins, and plays a crucial role in cholesterol transport, particularly  
 CC intracellular cholesterol trafficking in monocytes and fibroblasts, being  
 CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is  
 CC located on chromosome 9q31, and mutations in this gene are associated  
 CC with two genetic HDL (high density lipoprotein) deficiency disorders,  
 CC Tangier disease (TD) and familial HDL deficiency (FHA). These diseases  
 CC are distinguishable in that TD is an autosomal recessive disorder, while  
 CC FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good  
 CC cholesterol") in the blood correlate with a high risk of cardiovascular  
 CC disease, particularly coronary artery disease, but also cerebrovascular  
 CC disease, coronary restenosis, and peripheral vascular disease.  
 CC Conversely, a high level of HDL has protective effects against  
 CC cardiovascular disease. The invention provides genetic constructs and  
 CC transgenic cells and non-human animals comprising human ABC1 nucleic  
 CC acids, and methods of genetherapy for the treatment or prevention of  
 CC cardiovascular disease comprising the administration of an expression  
 CC vector encoding ABC1 or an active fragment thereof. The invention also  
 CC encompasses compounds which mimic ABC1 activity, compounds which  
 CC stimulate ABC1 expression and methods of screening for such compounds. It  
 CC further relates to methods for determining whether a patient has an  
 CC increased risk for cardiovascular disease due to polymorphisms in the  
 CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat or  
 CC prevent cardiovascular disease, especially coronary artery disease,  
 CC cerebrovascular disease, coronary restenosis or peripheral vascular  
 CC disease. They may also be used in the treatment of diseases associated  
 CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick  
 CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.  
 CC The invention specifically excludes proteins with the exact amino acid  
 CC sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic

CC acid with the exact sequence as GenBank Accession No: AJ012376.1. The  
CC present sequence represents a mutant human ABC1 cholesterol transporter  
CC associated with an altered cholesterol level and therefore an altered  
CC risk of cardiovascular disease. Note: The present sequence is not shown  
CC in the specification, but is derived from the native human ABC1 shown on  
CC pages 152-157

XX

SQ Sequence 2261 AA;

Query Match 33.5%; Score 4240.5; DB 3; Length 2261;  
Best Local Similarity 39.8%; Pred. No. 2.5e-307;  
Matches 1000; Conservative 345; Mismatches 730; Indels 435; Gaps 61;

Qy 6 QLQLLWKNVTLKRRSPWVLAFEIFIPLVLFILLGLRQKKPTISVKEVPFYTAAPL TSA 65  
||:|||||:| :|| | |: || :| ||: :| | | : ||  
Db 6 QLRLLWKNLTFRRRQTCQLLLEVAWPLFIFILISVRLSYPPYEQHECHFPNKA-MPSA 64

Qy 66 GILPVMQSLCPDGQRDEFGFL-----QYANSTVTQLLERLDRVVEEGNLFDPARP 115  
| || :| : : | : : | | :| |:: |:  
Db 65 GTLPWVQGIICNANNPCFRYPTPGEAPGVGNFNKSIVARLFSDARRLL----LYSQKDT 120

Qy 116 SLGSELEALR--QHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSLP 173  
|: : || |:: |: :| | | || |||||  
Db 121 SMKDMRKVLRTLQQIKKSSSNLKLQDFLVDNETFSG-----FLYHNLSLP 165

Qy 174 NSTAQALLAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELL LAPA 233  
|| :| | | :| :| | | | : ||::  
Db 166 KSTVDKMLRADV----ILHKVFLQGYQLHLTS-LCNGS-----KSEEMI---- 204

Qy 234 LLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAELRNQLDV 293  
|| | | : : :| : | : ||: :|:  
Db 205 ---QL---GDQEVSEL CGLPREKLAAE-----RVLRSNMDI 235

Qy 294 AK-VSQQLGLDAPNGSDSSPQAPPPRRLQALLGDLLD-----AQKVLQDQDVLS 341  
| : : | :| | :| :| || | |:: :| |  
Db 236 LKPILRTLNSTSPFPSKELAE--TKTLHLSLGTLAQELFSMRSWSDMRQEV MFLT NVNS 293

Qy 342 ALALLLPQGACTGRTPGPPASGAGGAAN-----GTGAGAVMGPNATAEEGAPSAAALATP 396  
: : | : | | | : | | | | : ||  
Db 294 SSSSTQIYQAVSRIVCGHPEGGLKIKSLNWYEDNNYKALFGGNGTEEDAETFYDNSTTP 353

Qy 397 ---DTLQGQCSAFVQ--LWAGLQPILCGNNR TIEPEALRRGNMSSLGFTSKEQRNLGLLV 451  
| :: | : : :| |::| |  
Db 354 YCNDLMKNLESSPLSRIIWKALKPLIVG----- 381

Qy 452 HLMTSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVWLNISAEIRSFLEQGR LQQ 511  
|||| | : : : |::| : :| :| :| :| :  
Db 382 -----KILYTPDTPATRQVMAEVNKT FQELAVFHDLEGMWEELSPKIWTFMENSQEMD 434

Qy 512 HLRWL-----QQYVAELRLHPE---ALNLSLDELPPALRQDNFS 547  
:| | | || | ||| :| | : | :|  
Db 435 LVRMLLDSRDNDHFWEQQLDGLDWT AQDIVAFLAKHPEDVQSSNGSVYTWREAFNETN-- 492

Qy 548 LPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQDNVTVFAS 607  
| : || :|| |:: : | :| :| | :|  
Db 493 -----QAIRTIS-----RFMECVNLNKLEPIATEVWLINKSME--LLDERKFWAG 535

Qy	608	VIFQTRKDG--LPPHVHYKIRQNSSFTEKTNERRAYWRPGPNTG---GRFYFLYGFVW	662
Db	536	IVFTGITPGSIELPHHVKYKIRMDIDNVERTNKIKDGYWDPGPRADPFEDMRVWGGFAY	595
Qy	663	IQDMMERAIIDTFVGHVDVVEPGSYVQMFPYPCYTRDDFLFVIEHMMPLCMVISWVYSVAM	722
Db	596	LQDVVEQAIIRVLTGTE-KKTGVYMQQMPYPCYVDDIFLRVMSRSMPLFMTLAWIYSVAV	654
Qy	723	TIQHIVAEKEHRLKEVMKTMGLNNAVHVVAVFITGTFVQLSISVTALTALIKYQVLMHSH	782
Db	655	IIKGIVYEKEARLKETMRIMGLDNSILWFSWFISSLIPLLVSAGLLVVILKLGNNLPYSD	714
Qy	783	VVIIWFLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHDK	842
Db	715	PSVVFVFLSVFAVVTILQCFLISTLFSRANLAAACGGIIYFTLYLPYVLC-----VAWQD	769
Qy	843	ITAFE-KCIASLMSTTAFLGLSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTMLM	901
Db	770	YVGFTLKI FASLLSPVAFGFGCEYFALFEEQGI GVQWDNLFESPVEEDGFNLTSVSMML	829
Qy	902	VDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSVM	961
Db	830	FDTFLYGVMTWYIEAVFPQYGI PRPWYFPCTKSYWFG-----ESDEKSHPGSNQKRIS--	884
Qy	962	EEDQACAMESRRFEETRGMEEPTHLPVVCVDKLTKVYKDDKKLALNKLNLNLYENQVV	1021
Db	885	---EIC-----MEEEPTHCLKGVSIQNLVKVYRDGMKVAVDGLALNFYEGQIT	929
Qy	1022	SFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLT	1081
Db	930	SFLGHNGAGKTTTMSILTGLFPPTSGTAYILGKDIRSEMSTIRQNLGVCPQHNVLFDMLT	989
Qy	1082	VEEHLWFYSRLKSMAQEEIRREMDKMIEDLEL-SNKRHSLVQTLSGGMKRKLSVAIAFVG	1140
Db	990	VEEHIWFYARLKLSEKHVKAEMEQLMALDVGLPSSSKLKS KTSQLSGGMQRKLSVALAFVG	1049
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Db	1050	GSKVVILDEPTAGVDPYSRRGIWELLKYRQGRTIILSTHHMDEADVLDRIAIISHGKL	1109
Qy	1201	KCCGSPLFLKGTYG DYRLTLVKRPAEPG-----GPQEPGLAS	1238
Db	1110	CCVGSSLFLKNQLGTGYLTLVKKDVESLSSCRNSSSTVSYLKKEDSVSQSSSDAGLGS	1169
Qy	1239	SPPGRAPLSSCSELQVSQFIRKHAVASCLLVSDTSTELSYILPSEAAKKGA FERLFQHLER	1298
Db	1170	DHDSDTLTIDVS--AISNLIRKHVSEARLVEDIGHELTYVLPYEAKEGAFVELFHEIDD	1227
Qy	1299	SLDALHLSSFGLMDTTLEEVFLKVSEEDQSLENSEADV KESRKDVLPGAEGPASGEGHAG	1358
Db	1228	RLSDLGISSYGISETTLEEIFLKVAEE-----SGVDA-ETSDGTLP-----	1267
Qy	1359	NLARCSELTQSQASLQSASSVGSARGDEGAGYTDVYGDYRPLF-DNPQDPD--NVSLQEV	1415
Db	1268	-----ARRNRR-FGDKQSCLRPFTEDDAADPNDSIDIPESR	1303
Qy	1416	EAEALSRV-GQGSRKLDGGWLKVROFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMT	1474



Db	1304	ETDLLSGMDGKGSYQVKGWKLTTQQQFVALLWKRLLIARRSRKGFFAQIVLPAVFCIALV	1363
Qy	1475	VALSVPEIGDLPPLVLSPSQYH-NYTQPRGNFIPYANEERREYRLRLSPDASPQQLVSTF	1533
Db	1364	FSLIVPPFGKYPSLELQPMWYNEQYT-----FVSNDAPF-----DTGTLELLNAL	1408
Qy	1534	RLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPP	1593
Db	1409	TKDPGFGTRCM-----EGNPI-----	1424
Qy	1594	PSPAPSDSPASPDEDLQAWNVSIPPTAGPEMWTSAFSLPRLVREPVR-----C	1641
Db	1425	-----PD-----TPCQAGEEWTAP-VPQTIMDLFQNGNWTMQNPSAC	1463
Qy	1642	TCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILTDTGHNVSLEYLLFTSDRF-----	1690
Db	1464	QCSSDKIKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLKTYVQIIAKSLKN	1523
Qy	1691	----RLHRYGATFG--NVLKSIPASFGTRAPPMVRK-----	1721
Db	1524	KIWNNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKHLKLAKDSSADRFLNSLGRFMTG	1583
Qy	1722	IAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSAS	1781
Db	1584	LDTRNNVKVWFNNKGWHAISSFLNVINAILRANLQKGE-NPSHYGITAFNHPNLTKQQ	1642
Qy	1782	LS-LDYLLQGTDVVIAIFIIVAMSFVPASFVVLVAEKSTKAKHLQFVSGCNPIIYWLAN	1840
Db	1643	LSEVALMTTSVDVLVSICVIFAMSFVPASFVFLIQERVSKAKHLQFISGVKPVYWLSN	1702
Qy	1841	YVWDMNLNYLVPATCCVIIIFVFDLPAYTSPTNFPVAVLSLFLLYGWSITPIMYPASFWEV	1900
Db	1703	FVWDMCNVVPATLVIIIFICFQQKSYVSSTNLVPLALLLLLYGWSITPLMPYASFVFKI	1762
Qy	1901	PSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLME	1960
Db	1763	PSTAYVVLTSVNLFIGINGSVATFVLELFT-DNKLNNINDILKSVLIFPHFCLGRGLID	1821
Qy	1961	MAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQR	2020
Db	1822	MVKNQAMADALERFGE-NRFVSPLSWDLVGRNLFAMAVEGVVFFLITVLIQYRFFIRPRP	1880
Qy	2021	MPVSTKPVED-DVDVASERQVRGRGDADNDMVKIENTKVYKSRKIGRILAVDRCLGVR	2079
Db	1881	VNAKLSPLNDEDEDVRRERQRILDGGGQNDILEIKELTKIYRK---RKPAVDRICVGIP	1937
Qy	2080	PGECEGGLGVNGAGKTSTFKMLTGDESTTGGEAFVNGHSLVKELLQVQQSLGYCPQCDAL	2139
Db	1938	PGECEGGLGVNGAGKSSTFKMLTGDTTVTRGDAFLNKNLSILNIHEVHQNMGYCPQFQDAI	1997
Qy	2140	FDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAI	2199
Db	1998	TELLTGREHVEFFALLRGVPEKEVGKGEWAIKRLGLVKYGEKYAGNYSGGNKRKLSTAM	2057
Qy	2200	ALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIM	2259

Db 2058 ALIGGPPVFLDEPTTGMDPKARRFLWNCALSVVKEGRSVVLTSHSMEECEALCTRMAIM 2117  
 Qy 2260 VNGLRCLGSIQHLKNRFGDGYMITVR-TKSSQSVKDVVREFFNRNFPEAMLERHHTKVQ 2318  
 |||| ||||:||||||| | | | : :| | | | || :|||:| :|  
 Db 2118 VNGRFRCLGSVQHLKNRFGDGYTIVVRIAGSNPDLKPVQDFFGLAFPGSVLKEKHRNMLQ 2177  
 Qy 2319 YQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFVNFAKKQSDN 2368  
 ||| | |||:| | : | | ||||| ||||| |||:  
 Db 2178 YQLPSSLSSLARIFSILSQSKKRLHIEDYSVSQTTLDQVFNFAKDQSD 2227

Search completed: September 1, 2004, 10:52:37  
 Job time : 223 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2004, 10:49:12 ; Search time 45 Seconds  
(without alignments)  
2794.686 Million cell updates/sec

Title: US-10-088-467-2  
Perfect score: 12668  
Sequence: 1 MGFLHQLQLLLWKNVTLKRR.....GLISFEEERAQLSFNTDTLC 2436

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	%		Query			ID	Description
	Score	Match	Length	DB			
1	7046	55.6	1457	3	US-08-665-259-27		Sequence 27, Appl
2	7046	55.6	1457	3	US-08-762-500-27		Sequence 27, Appl
3	4240.5	33.5	2261	4	US-09-526-193A-1		Sequence 1, Appli
4	3173.5	25.1	1375	3	US-08-665-259-26		Sequence 26, Appl
5	3173.5	25.1	1375	3	US-08-762-500-26		Sequence 26, Appl
6	2622	20.7	1684	3	US-08-665-259-25		Sequence 25, Appl
7	2622	20.7	1684	3	US-08-762-500-25		Sequence 25, Appl
8	2622	20.7	1704	3	US-08-762-500-75		Sequence 75, Appl
9	363.5	2.9	589	4	US-09-328-352-7592		Sequence 7592, Ap
10	360	2.8	317	4	US-09-489-039A-10626		Sequence 10626, A
11	359.5	2.8	345	4	US-09-252-991A-31957		Sequence 31957, A

12	343.5	2.7	532	4	US-09-543-681A-4646	Sequence 4646, Ap
13	341.5	2.7	788	4	US-09-252-991A-28171	Sequence 28171, A
14	339.5	2.7	315	4	US-09-134-000C-6449	Sequence 6449, Ap
15	338.5	2.7	607	4	US-09-252-991A-18351	Sequence 18351, A
16	337.5	2.7	315	4	US-09-328-352-4388	Sequence 4388, Ap
17	336	2.7	309	4	US-09-252-991A-21204	Sequence 21204, A
18	334	2.6	335	4	US-09-252-991A-20837	Sequence 20837, A
19	333	2.6	588	4	US-09-489-039A-13579	Sequence 13579, A
20	330.5	2.6	922	4	US-09-489-039A-8938	Sequence 8938, Ap
21	330	2.6	594	4	US-09-543-681A-5528	Sequence 5528, Ap
22	328	2.6	332	4	US-09-107-532A-3752	Sequence 3752, Ap
23	327.5	2.6	304	4	US-09-107-532A-5424	Sequence 5424, Ap
24	327	2.6	291	4	US-09-107-532A-4205	Sequence 4205, Ap
25	327	2.6	929	4	US-09-252-991A-22946	Sequence 22946, A
26	326.5	2.6	323	4	US-09-489-039A-12496	Sequence 12496, A
27	322.5	2.5	322	4	US-09-107-532A-4662	Sequence 4662, Ap
28	316.5	2.5	316	4	US-09-543-681A-6184	Sequence 6184, Ap
29	315.5	2.5	1280	2	US-08-583-276-19	Sequence 19, Appl
30	315	2.5	1279	4	US-09-672-810-6	Sequence 6, Appli
31	314.5	2.5	1280	4	US-09-767-594-2	Sequence 2, Appli
32	314.5	2.5	1280	4	US-09-672-810-5	Sequence 5, Appli
33	314.5	2.5	1280	6	5206352-4	Patent No. 5206352
34	314	2.5	233	4	US-09-627-376-12	Sequence 12, Appl
35	312.5	2.5	1280	4	US-09-672-810-2	Sequence 2, Appli
36	312.5	2.5	1283	4	US-09-672-810-4	Sequence 4, Appli
37	312	2.5	1279	2	US-08-784-649A-2	Sequence 2, Appli
38	309.5	2.4	402	4	US-09-107-532A-5360	Sequence 5360, Ap
39	308.5	2.4	391	4	US-09-252-991A-20275	Sequence 20275, A
40	307.5	2.4	1280	2	US-08-752-447-2	Sequence 2, Appli
41	307.5	2.4	1280	4	US-09-316-167-2	Sequence 2, Appli
42	307.5	2.4	1280	4	US-09-397-233-2	Sequence 2, Appli
43	304.5	2.4	231	4	US-09-134-001C-3824	Sequence 3824, Ap
44	303.5	2.4	243	4	US-09-543-681A-5911	Sequence 5911, Ap
45	301.5	2.4	254	4	US-09-107-532A-4983	Sequence 4983, Ap

#### ALIGNMENTS

##### RESULT 1

US-08-665-259-27

; Sequence 27, Application US/08665259

; Patent No. 6028173

##### ; GENERAL INFORMATION:

; APPLICANT: Landes, Gregory M.

; APPLICANT: Burn, Timothy C.

; APPLICANT: Connors, Timothy D.

; APPLICANT: Dackowski, William R.

; APPLICANT: Van Raay, Terence J.

; APPLICANT: Klinger, Katherine W.

; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,

; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME

; NUMBER OF SEQUENCES: 73

##### ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENZYME CORPORATION

; STREET: One Mountain Road

; CITY: Framingham



Db	361	RKDVLPGAEGLTAVGGQAGNLARCSELAQSQASLQSSASSVGSARGEEGTGYSDGYGDYRP	420
Qy	1400	LFDNPQDPDNVSLQEVEAEALS RVGQGS RKLDGGWLKVRQFHGLLVKR FHCARRNSKALF	1459
Db	421	LFDNLQDPDNVSLQEAE MEALAQVGQGS RKLEGW LKMRQFHGLLVKR FHCARRNSKALC	480
Qy	1460	SQILLPAFFVCVAMTVALS VPEIGDL PPLVLSPSQYHNYTQPRGNFI PYANEERREYRLR	1519
Db	481	SQILLPAFFVCVAMTVALS VPEIGDL PPLVLSPSQYHNYTQPRGNFI PYANEERQEYRLR	540
Qy	1520	LSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFDSMCLES	1579
Db	541	LSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPMNLNLSSGESRLLAARFFDSMCLES	600
Qy	1580	FTQGLPLSNFVPPPPSPAPSDSPASPD ED-LQAWNVS LPPTAGPEMWT SAPSLPRLVREP	1638
Db	601	FTQGLPLSNFVPPPPSPAPSDSPVXPDEDSLQAWNMSLPPTAGPETWTSAPSLPRLVHEP	660
Qy	1639	VRCTCSAQGTGFSCPSVGGHPPQMRVVTGDI LT DITGHN VSEYLLFTSDRFLHRYGAI	1698
Db	661	VRCTCSAQGTGFSCPSVGGHPPQMRVVTGDI LT DITGHN VSEYLLFTSDRFLHRYGAI	720
Qy	1699	TFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPK	1758
Db	721	TFGNVQKSIPASFGARVPPMVRKIAVRRAQVLYNNKGYHSMPTYLNSLNNAILRANLPK	780
Qy	1759	SKGNPAAYGITVTNHPMNKTSASLSLDYLLQGT DVVIAIFII VAMSFVPASFVVFLVAEK	1818
Db	781	SKGNPAAYXITVTNHPMNKTSASLSLDYLLQGT DVVIAIFII VAMSFVPASFVVFLVAEK	840
Qy	1819	STKAKHLQFVSGCNPIIYWLANYVWDM LNYLV PATCCVII LFVFDLPAYTSPTNFP AVL S	1878
Db	841	STKAKHLQFVSGCNPVIYWLANYVWDM LNYLV PATCCVII LFVFDLPAYTSPTNFP AVL S	900
Qy	1879	LFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKV V	1938
Db	901	LFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKV V	960
Qy	1939	NSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAV	1998
Db	961	NSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMTV	1020
Qy	1999	EGVVGFLLTIMCQYNFLRRPQRMVSTKPV EDDVDVASERQ RVLRGDADNDMVKIENLTK	2058
Db	1021	EGFVGFFLTIMCQYNFLRQPQRLPVSTKPV EDDVDVASERQ RVLRGDADNDMVKIENLTK	1080
Qy	2059	VYKSRKIGRILAVDRLCLGV-RPGE CFGL LGVNGAGKTSTFKMLTGDESTTGGEAFVNGH	2117
Db	1081	VYKSRKIGRILAVDRLCLGVCVPGE CFGL LGVNGAGKTSTFKMLTGDESTTGGEAFVNGH	1140
Qy	2118	SVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELT	2177
Db	1141	SVLKDLLQVQQSLGYCPQFDPVDELTA REHLQLYTRLRCIPWKDEAQVVKWALEKLELT	1200
Qy	2178	KYADKPAGTYSGGNKRKLSTAIALIGYPAFI FLDEPTTGMDPKARRFLWNLILD LIKTGR	2237

Db 1201 KYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIKTGR 1260

Qy 2238 SVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVVRTKSSQSVKDVV 2297  
 |||||:||||

Db 1261 SVVLTSHSMEECEALCTRLAIMVNGRLHCLGSIQHLKNRFGDGYMITVVRTKSSQNVKDVV 1320

Qy 2298 RFFNRNFPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNV 2357  
 ||||| : : |||||

Db 1321 RFFNRNFPEAHAQGKTPYKVQYQLKSEHISLAQVFSKMEQVVGVLGIEDYSVSQTTLDNV 1380

Qy 2358 FVNFAKKQSDNLEQQETEPFSALQSPLGCLLSLLRPRAPTELRALVADEPEDLDTEDEG 2417  
 |||||:|||| ||| |||||

Db 1381 FVNFAKKQSDNVEQQEAE-PSSLPSPLG-LLSLLRPRAPTELRALVADEPEDLDTEDEG 1438

Qy 2418 LISFEEERAQLSFNTDTLC 2436  
 |||||

Db 1439 LISFEEERAQLSFNTDTLC 1457

# RESULT 2

US-08-762-500-27

; Sequence 27, Application US/08762500

; Patent No. 6030806

## ; GENERAL INFORMATION:

; APPLICANT: Landes, Gregory M.

; APPLICANT: Burn, Timothy C.

; APPLICANT: Connors, Timothy D.

; APPLICANT: Dackowski, William R.

; APPLICANT: Van Raay, Terence J.

; APPLICANT: Klinger, Katherine W.

; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,

; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME

; NUMBER OF SEQUENCES: 83

## ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENZYME CORPORATION

; STREET: One Mountain Road

; CITY: Framingham

; STATE: Massachusetts

; COUNTRY: United States of America

; ZIP: 01701

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

## ; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/762,500

; FILING DATE: 09-DEC-1996

; CLASSIFICATION: 435

## ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/665,259

; FILING DATE: 17-JUN-1996

## ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US96/10469

; FILING DATE: 17-JUN-1996

## ; ATTORNEY/AGENT INFORMATION:

; NAME: Dugan, Deborah A.

; REGISTRATION NUMBER: 37,315  
 ; REFERENCE/DOCKET NUMBER: IG5-9.3  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (508) 872-8400  
 ; TELEFAX: (508) 872-5415  
 ; INFORMATION FOR SEQ ID NO: 27:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1457 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 US-08-762-500-27

Query Match 55.6%; Score 7046; DB 3; Length 1457;  
 Best Local Similarity 94.2%; Pred. No. 0;  
 Matches 1374; Conservative 22; Mismatches 59; Indels 4; Gaps 4;

Qy	980	MEEEPHTLPLVVCVDKLT	KVYKDDKKLALNKL	SLNLYENQVVSFLGHNGAGKTTTMSILT	1039
Db	1	MEEEPHTLPLVVCVDKLT	KVYKNDKKLALNKL	SLNLYENQVVSFLGHNGAGKTTTMSILT	60
Qy	1040	GLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR	LTVEEHLWFYSRLKSMAQEE		1099
Db	61	GLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR	LTVEEHLWFYSRLKSMAQEE		120
Qy	1100	IRREMDKMIEDLELSNKRHSLVQTL	SGGMKRKLSVAIAFVGGSRAIILDEPTAGVDPYAR		1159
Db	121	IRKETDKMIEDLELSNKRHSLVQTL	SGGMKRKLSVAIAFVGGSRAIILDEPTAGVDPYAR		180
Qy	1160	RAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAII	SHGKLKCCGSPLFLKGTYGDGYRL		1219
Db	181	RAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAII	SHGKLKCCGSPLFLKGAYXDGYRL		240
Qy	1220	TLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHVASC	LLVSDTSTELSYIL		1279
Db	241	TLVKQPAEPGTSQEPGLASSPSGCPRLSSCSEPQVSQFIRKHVASS	LLVSDTSTELSYIL		300
Qy	1280	PSEAAKKGAFERLFQHLERSLDALHLSSFG	LMDTTLEEVLKVSEEDQSL	ENSEADVKE	1339
Db	301	PSEAVKKGAFERLFQQLHSLDALHLSSFG	LMDTTLEEVLKVSEEDQSL	ENSEADVKE	360
Qy	1340	RKDVLPGAEGPASGEGHAGNLARCSELTQSQASLQSASSVGSARG	DEGAGYTDVYGDYRP		1399
			:		
Db	361	RKDVLPGAEGLTAVGGQAGNLARCSELAQSQASLQSASSVGSARGE	EGTGYS	SDGYGDYRP	420
Qy	1400	LFDNPQDPDNVSLQEVEAEALS	RVGQGSRKLDGGWLKVRQFHGLLVKRFHCARRNSKALF		1459
Db	421	LFDNLQDPDNVSLQEAEAEALAQVGQGSRKLEGWWLKM	RQFHGLLVKRFHCARRNSKALC		480
Qy	1460	SQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQPRGNFI	PYANEERREYRLR		1519
Db	481	SQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQPRGNFI	PYANEERQEYRLR		540
Qy	1520	LSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLN	LSSGESRLAARFFDSMCLES		1579
Db	541	LSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLN	LSSGESRLAARFFDSMCLES		600



Qy	1580	FTQGLPLSNFVPPPPSPAPSDSPASPD	ED-LQAWNVS	LPPTAGPEM	WTSAPSLPRLVREP	1638
				:		
Db	601	FTQGLPLSNFVPPPPSPAPSDSPVX	PD	EDSLQAWN	MSLPPTAGPETWTSAPSLPRLVHEP	660
Qy	1639	VRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDIL	TDITGHN	VSEYLLFTSDR	FRLHRYGAI	1698
Db	661	VRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDIL	TDITGHN	VSEYLLFTSDR	FRLHRYGAI	720
Qy	1699	TFGNVLKSIPASFGTRAPPMVRKIAVRR	AAQVFYNN	KGYS	HMPTYLNSLNNAILRANLPK	1758
Db	721	TFGNVQKSIPASFGARVPPMVRKIAVRR	VAQVLYNN	KGYS	HMPTYLNSLNNAILRANLPK	780
Qy	1759	SKGNPAAYGITVTNHPMNKTSASLSLDYLLQGT	DVVIAIFII	VAMSFVP	PASFVFLVAEK	1818
Db	781	SKGNPAAYXITVTNHPMNKTSASLSLDYLLQGT	DVVIAIFII	VAMSFVP	PASFVFLVAEK	840
Qy	1819	STKAKHLQFVSGCNP	IIYWLANYVW	DMLNYL	VPATCCVILFVFDLPAYTSPTNFP	PAVLS 1878
Db	841	STKAKHLQFVSGCNP	VIYWLANYVW	DMLNYL	VPATCCVILFVFDLPAYTSPTNFP	PAVLS 900
Qy	1879	LFLLYGWSITPIMYPASFWF	VEVPSSAYVFL	VINL	FIGITATVATFLLQLFEHDKDLKVV	1938
Db	901	LFLLYGWSITPIMYPASFWF	VEVPSSAYVFL	VINL	FIGITATVATFLLQLFEHDKDLKVV	960
Qy	1939	NSYLKSCFLIFPNYNLGHGLMEMAYNEYINEY	YAKIGQF	DKMKSPFEW	DIVTRGLVAMAV	1998
Db	961	NSYLKSCFLIFPNYNLGHGLMEMAYNEYINEY	YAKIGQF	DKMKSPFEW	DIVTRGLVAMTV	1020
Qy	1999	EGVVGFLLTIMCQYNFLRRPQRM	PVSTKPVEDD	VDV	VASERQVR	LRGDADNDMVKIENLTK 2058
				:		
Db	1021	EGFVGFFLTIMCQYNFLRQ	RPQRLPVSTKPVEDD	VDV	VASERQVR	LRGDADNDMVKIENLTK 1080
Qy	2059	VYKSRKIGRILAVDRLCLGV-RP	GECFGLLG	VNGAGKTSTFKMLTGDESTTGGEAFVNGH	2117	
Db	1081	VYKSRKIGRILAVDRLCLGV	CVPGECFGLLG	VNGAGKTSTFKMLTGDESTTGGEAFVNGH	1140	
Qy	2118	SVLKELLQVQQSLGYCPQCDALFDELTAREHLQ	LYTRLRGISWKDEARVVKWALEKLELT	2177		
		:			:	
Db	1141	SVLKDLLQVQQSLGYCPQFDPVDEL	TAREHLQLYTRLRCIPWKDEAQVVKWALEKLELT	1200		
Qy	2178	KYADKPAGTYS	GGNKRKLSTAIALIGYP	AFIFLDEPTTGMDPKARRFLWN	LILDLIKTGR	2237
Db	1201	KYADKPAGTYS	GGNKRKLSTAIALIGYP	AFIFLDEPTTGMDPKARRFLWN	LILDLIKTGR	1260
Qy	2238	SVVLTSHSMEECEALCTR	LAIMVNGRLRCLGSIQHLKNRFGDGYMITV	RTKSSQSVKDVV	2297	
					:	
Db	1261	SVVLTSHSMEECEALCTR	LAIMVNGRLHCLGSIQHLKNRFGDGYMITV	RTKSSQNVKDVV	1320	
Qy	2298	RFFNRNFP	PEAM	LKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNV	2357	
			:	:		
Db	1321	RFFNRNFP	EAHAQ	GKTPYKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNV	1380	
Qy	2358	FVNFAKKQSDNLEQQ	ETEP	PSALQSPLGCLLSLLRPRSAPTEL	RALVADEPEDLDT	TEDEG 2417
			:			
Db	1381	FVNFAKKQSDNVEQQ	EAE-PSSLP	SPLG-LLSLLRPRPAPTEL	RALVADEPEDLDT	TEDEG 1438

Qy 2418 LISFEEERAQLSFNTDTLC 2436  
|||||||  
Db 1439 LISFEEERAQLSFNTDTLC 1457

RESULT 3

US-09-526-193A-1

; Sequence 1, Application US/09526193A  
; Patent No. 6617122  
; GENERAL INFORMATION:  
; APPLICANT: Hayden, Michael R.  
; APPLICANT: Brooks-Wilson, Angela R.  
; APPLICANT: Pimstone, Simon N.  
; TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING  
; TITLE OF INVENTION: CHOLESTEROL LEVELS  
; FILE REFERENCE: 50110/002005  
; CURRENT APPLICATION NUMBER: US/09/526,193A  
; CURRENT FILING DATE: 2000-03-15  
; PRIOR APPLICATION NUMBER: 60/124,702  
; PRIOR FILING DATE: 1999-03-15  
; PRIOR APPLICATION NUMBER: 60/138,048  
; PRIOR FILING DATE: 1999-06-08  
; PRIOR APPLICATION NUMBER: 60/139,600  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: 60/151,977  
; PRIOR FILING DATE: 1999-09-01  
; NUMBER OF SEQ ID NOS: 287  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2261  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-526-193A-1

Query Match 33.5%; Score 4240.5; DB 4; Length 2261;  
Best Local Similarity 39.8%; Pred. No. 0;  
Matches 1000; Conservative 345; Mismatches 730; Indels 435; Gaps 61;

Qy 6 QLQLLLWKNVTLKRRSPWVLAFEIFIFLVLFFILLGLRQKKPTISVKEVPFYTAAPL TSA 65  
||:|||||:| :|| | |: || :| ||: :| | | : ||  
Db 6 QLRLLLWKNLTFRRRQTCQLLLEVAWPLFIFLILISVRLSYPPYEQHECHFPNKA-MPSA 64  
Qy 66 GILPVMQSLCPDGQRDEFGL-----QYANSTVTQLLERLDRVVEEGNLFDPARP 115  
| || :| : : | : : | | :| | : : | :  
Db 65 GTLPWVQGIICNANNPCFRYPTPGEAPGVVGNFNKSIVARLFSDARRLL----LYSQKDT 120  
Qy 116 SLGSELEALR--QHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSLP 173  
|: : || | : : |: :| || | | || ||||  
Db 121 SMKDMRKVLRTLQQIKSSSNLKLQDFLVDNETFSG-----FLYHNLSLP 165  
Qy 174 NSTAQALLAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPA 233  
|| :| | | :| :| | | | : ||: :  
Db 166 KSTVDKMLRADV----ILHKVFLQGYQLHLTS-LCNGS-----KSEEMI---- 204  
Qy 234 LLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAELRNQLDV 293  
|| | | : : :| : | : ||: :| :  
Db 205 ---QL----GDQEVSELCPREKLAAAE-----RVLRSNMDI 235

Qy	294	AK-VSQQGLDAPNGSDSSPQAPPPRRLLQALLGDLDD-----AQKVLQDQDVL	341
Db	236	LKPILRTLNSTSPFPKELAE--TKTLLHSLGTLAQELFSMRSWSDMRQEVMTLVNS	293
Qy	342	ALALLLPQGACTGRTPGPPASGAGGAAN-----GTGAGAVMGPNATAEEGAPSAAALATP	396
Db	294	SSSSTQIYQAVSRIVCGHPEGGLKIKSLNWDYEDNNYKALFGGNGTEEDAETFYDNSTTP	353
Qy	397	---DTLQGQCSAFVQ---LWAGLQPILCGNNRTIEPEALRRGNMSSLGFTSKEQRNLGLLV	451
Db	354	YCNDLMKNLESSPLSRIIWKALKPLLVG-----	381
Qy	452	HLMTSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVWLNISAEIRSFLEQGRLQQ	511
Db	382	-----KILYTPDTPATROVMAEVNKTQELAVFHDLEGMWEELSPKIWTFMENSQEMD	434
Qy	512	HLRWL-----QQYVAELRLHPE---ALNLSLDELPPALRQDNFS	547
Db	435	LVRMLLDSRDNDHFWEQQLDGLDWTQAQDIVAFLAKHPEDVQSSNGSVYTWREAFNETN--	492
Qy	548	LPSGMALLQQLDITDNAACGWIQFMSKVSVDIFKGFPEDEESIVNYTLNQAYQDNVTVFAS	607
Db	493	-----QAIRTIS-----RFMECVNLNKLEPIATEVWLINKSME--LLDERKEFWAG	535
Qy	608	VIFQTRKDG--LPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTG---GRFYFLYGFVW	662
Db	536	IVFTGITPGSIELPHHVYKIRMDIDNVERTNKIKDGYWDPGPRADPFEDMRYVWGGFAY	595
Qy	663	IQDMMERAIIDTFVGHVVEPGSYVQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSVAM	722
Db	596	LQDVVEQAIIRVLTGTE-KKTGVYMQMPYPCYVDDIFLVRMSRSMPLFMTLAWIYSVAV	654
Qy	723	TIQHIVAEKEHRLKEVMKTMGLNNVHVVAVFITGFVQLSISVTALTALIKYQVLMHSH	782
Db	655	IIKGIVYEKEARLKETMRIMGLDNSILWFSWFISLIPLLVSAGLLVILKGLNLLPYSD	714
Qy	783	VVIIWFLFVAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHDK	842
Db	715	PSVVFVFLSVFAVVTILQCFLISTLFSRANLAAACGGIIYFTLYLPYVLC-----VAWQD	769
Qy	843	ITAFE-KCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTMLM	901
Db	770	YVGFTLKIFASLLSPVAFGFGCEYFALFEEQGIGVQWDNLFESPVEEDGFNLTSVSMML	829
Qy	902	VDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSVM	961
Db	830	FDTFLYGVMTWYIEAVFPGQYGIIPRWYFPCTKSYWFG--ESDEKSHPGSNQKRIS--	884
Qy	962	EEDQACAMESRRFEETRGMEEPTHLPLVVCVDKLTKVYKDDKKLALNKLNLNLYENQVV	1021
Db	885	---EIC-----MEEPTHLKLGVSIGNLVKVYRDGMKVAVDGLALNFYEGQIT	929
Qy	1022	SFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRIT	1081
Db	930	SFLGHNGAGKTTTMSILTGLFPPTSGTAYILGKDIRSEMSTIRONLGVCPHNVLFDMIT	989

Qy	1082	VEEHLWFYSRKLSMAQEEIRREMDKMIEDLEL-SNKRHSLVQTLSSGGMKRKLSVAIAFVG	1140
Db	990	VEEHIWFYARLKGLSEKHVKAEMEQLMALDVGLPSSKLKSKTSQLSGGMQRKLSVALAFVG	1049
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Db	1050	GSKVVILDEPTAGVDPYSRRGIWELLKRYQGRTIILSTHHMDEADVLGDRIAIISHGKL	1109
Qy	1201	KCCGSPLFLKGTGYDGYRLTLVKRPAEPG-----GPQEPGLAS	1238
Db	1110	CCVGSSLFLKNQLGTGYLTTLVKKDVESSLSSCRNSSSTVSYLKKEDSVSQSSSDAGLGS	1169
Qy	1239	SPPGRAPLSSCSELQVSQFIRKHAVASCLLVSDTSTELSYILPSEAAKKGAFERLFQHLER	1298
Db	1170	DHESDTLTIDVS--AISNLIRKHVSEARLVEDIGHELTIVLPYEAAKEGAFVELFHEIDD	1227
Qy	1299	SIDALHLSSFGMLDITLLEEVFLKVSEEDQSLNSEADVKESRKDVLPGAEGPASGEGHAG	1358
Db	1228	RLSDLGISSYGISETTLEEIFLKVAEE-----SGVDA-ETSDGTLP-----	1267
Qy	1359	NLARCSELTQSQASLQSASSVGSARGDEGAGYTDVYGDYRPLF-DNPQDPD--NVSLQEV	1415
Db	1268	-----ARRNRRRA-FGDKQSCLRPFTEDDAADPNDSIDIPESR	1303
Qy	1416	EAEALSRV-GQGSRKLDGGWLKVRQFHGLLVKRFHCARRNSKALFSQILLPAFFVVCVAMT	1474
Db	1304	ETDLLSGMDGKGSYQVKGWKLTTQQQFVALLWKRLLIARRSRKGFFAQIVLPAVFVCIALV	1363
Qy	1475	VALSVPEIGDLPLPLVLSPSQYH-NYTQPRGNFIPIYANEERREYRLRLSPDASPQQLVSTF	1533
Db	1364	FSLIVPPFGKYPSLELQPMWYNEQYT-----FVSNDAPF-----DTGTLELLNAL	1408
Qy	1534	RLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPP	1593
Db	1409	TKDPGFGTRCM-----EGNPI-----	1424
Qy	1594	PSPAPSDSPASPDEDLQAWNVSLLPPTAGPEMWTSAFSLPRLVREPVR-----C	1641
Db	1425	-----PD-----TPCQAGEEWTAP-VPQTIMDLFQNGNWTMQNPSAC	1463
Qy	1642	TCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILTDTGHNVSLEYLLFTSDRF-----	1690
Db	1464	QCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKN	1523
Qy	1691	----RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK-----	1721
Db	1524	KIWNNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKHKLAKDSSADRFLNSLGRFMTG	1583
Qy	1722	IARRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSAS	1781
Db	1584	LDTRNNVKVWFNNKGWHAISSFLNVINNAILRANLQKGE-NPSHYGITAFNHPNLTKQQ	1642
Qy	1782	LS-LDYLLQGTDVVIAIFIIVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLAN	1840
Db	1643	LSEVALMTTSDVVLVSICVIFAMSFVPASFVFLVQERVSKAKHLQFISGVKPVIIYWLSN	1702
Qy	1841	YVWDMNLNLYLPATCCVILFVFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFVFEV	1900

```

      :||| ||:|||| :|| | :| || | : | |||||:|||| |::
Db      1703 FVWDMCNYVVPATLVIIIFICFQQKSYVSSTNLPVLALLLLLYGWSITPLMYPASFVKI 1762

Qy      1901 PSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLME 1960
      ||:|| | :|||| | :|||:|:| | | :| || | ||||:| | ||::
Db      1763 PSTAYVVLTSVNLFINGSVATFVLELFT-DNKLNNINDILKSVFLIFPHFCLGRGLID 1821

Qy      1961 MAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQR 2020
      | |: : : : |: :: || ||:| | | ||||| ||:|:| | | ||:
Db      1822 MVKNQAMADALERFGE-NRFVSPLSWDLVGRNLFAMAVEGVVFFLITVLIQYRFFIRPRP 1880

Qy      2021 MPVSTKPVED-DVDVASERQVRVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVR 2079
      : | : | | || |||:| | ||::|: |||:|: : | |||:|:|:
Db      1881 VNAKLSPLNDEDEDVRRERQRILDGGGQNDILEIKELTKIYRRK---RKPAVDRICVGIP 1937

Qy      2080 PGEFCGLLGVNGAGKTSTFKMLTGDESTTGGEAFVNGHSLVKELLQVQQSLGYCPQCDAL 2139
      ||||| |||||:||||| | :| |:|:| |:| : :| |:| ||| ||:
Db      1938 PGEFCGLLGVNGAGKSSTFKMLTGDTTVTRGDAFLNKNLSILSNIHEVHQNMGYCPQFDAI 1997

Qy      2140 FDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAI 2199
      : || |||: : |||: |: :| :|: || | || :| || ||||| |||:
Db      1998 TELLTGREHVEFFALLRGVPEKEVGKVGWAIRKGLVKYGEKYAGNYSGGNKRKLSTAM 2057

Qy      2200 ALIGYPAFIFLDEPTTGMDPKARRFLWNLIIDLIKTGRSVVLTSHSMEECEALCTRLAIM 2259
      |||| | :||||| ||||| | :| ||||| |||||:| |
Db      2058 ALIGGPPVVFLDEPTTGMDPKARRFLWNCALSVVKEGRSVVLTSHSMEECEALCTRMAM 2117

Qy      2260 VNGLRLCLGSIQHLKNRFGDGYMITVR-TKSSQSVKDVVRFFNRNFP EAMLERHHTKVQ 2318
      |||| ||||:||||| | || |: :| | || || :|||:| :|
Db      2118 VNGRFRCLGQSVQHLKNRFGDGYTIVVRIAGSNPDLKPVQDFFGLAFPGSVLKEKHRNMLQ 2177

Qy      2319 YQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFVNFAKKQSDN 2368
      ||| | |||:| | :| | ||||| ||||| |||:
Db      2178 YQLPSSLSSLARIFSILSQSKRLHIEDYSVSQTTLDQVFNFAKDQSD 2227

```

RESULT 4

US-08-665-259-26

; Sequence 26, Application US/08665259

; Patent No. 6028173

; GENERAL INFORMATION:

; APPLICANT: Landes, Gregory M.

; APPLICANT: Burn, Timothy C.

; APPLICANT: Connors, Timothy D.

; APPLICANT: Dackowski, William R.

; APPLICANT: Van Raay, Terence J.

; APPLICANT: Klinger, Katherine W.

; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,

; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME

; NUMBER OF SEQUENCES: 73

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENZYME CORPORATION

; STREET: One Mountain Road

; CITY: Framingham

; STATE: Massachusetts

; COUNTRY: United States of America

; ZIP: 01701

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,259
; FILING DATE: 17-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IG5-9.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1375 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-665-259-26

```

```

Query Match          25.1%; Score 3173.5; DB 3; Length 1375;
Best Local Similarity 46.5%; Pred. No. 5.4e-274;
Matches 683; Conservative 205; Mismatches 372; Indels 209; Gaps 30;

```

```

Qy      980 MEEEPHTLPLVVCVDKLTKVYKDDKKLALNKLSTLNLYENQVVSFLGHNGAGKTTTMSILT 1039
      ||||| | | : | ||:| |:|:| |:| | | : ||||| |||||
Db      2 MEEEPHTLRLGVSIQNLVKVYRDGMKVAVDGLALNFYEGQITSFLGHNGAGKTTTMSILT 61

Qy     1040 GLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSRLKSMAQEE 1099
      ||||| |:| | | ||:| | |:|:| ||||| ||||| |:|:| | | :|:|
Db      62 GLFPPTSGTAYILGKDIRSEMSSIRQNLGVCPQHNVLFDMLTVEEHIWFYARLKGLSEKH 121

Qy     1100 IRREMDKMIEDLEL-SNKRHSLVQTLSSGGMKRKLSVAIAFVGGSRAIILDEPTAGVDPYA 1158
      :: ||:| | | : | | | |||:|||||:|||||: :||||| |||||
Db      122 VKAEMEQMALDVGLPPSKLKSQTSQSSGGMQRKLSVALAFVGGSKVILDEPTAGVDPYS 181

Qy     1159 RRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPFLFKGTYG DGYR 1218
      || ||:|:| | | |||:|||||:||||| || || ||| | ||
Db      182 RRGIWELLKYRQGRTIILSTHHMDEADILGDRIAIISHGKLCCVGSSLFLKNQLGTGY 241

Qy     1219 LTLVKRPAEPG-----GPQEPGLASSPPGRAPLSSCSELQVSQ 1256
      ||||: | : || | | :|
Db      242 LTLVKKDVESLSSCRNSSSTVSLCKKEDSVSQSSSDAGLGSDHESDTLTIDVS--AISN 299

Qy     1257 FIRKHVASCLLVSDTSTELSYILPSEAAKKGAFERLFQHLESLDALHLSFGLMDTTLE 1316
      ||||: || | ||:|:| |||:| | || :| | | :|:| :| | |
Db      300 LIRKHVSEARLVEDIGHELTYPYEAKEGAFVELFHEIDRLSDLGISYGISETTLE 359

Qy     1317 EVFLKVSEEDQSLENSEADVKE SRKDVLPGAEGPASGEGHAGNLARCSELTQSQASLQSA 1376
      |:| |:| | | | : | |
Db      360 EIFLKVAEE-----SGVDA-ETSDGTLP----- 381

```

Qy	1377	SSVGSARGDEGAGYTDVYGDYRPLF-----DNPQDPD--NVSLQEVEAEALSRV-GQGSR	1428
Db	382	-----ARRNRA-----FGDKQSCSLHPFTEDDAVDPNDSIDPESRETDLLSGMDGKGSY	431
Qy	1429	KLDGGWLKVRQFHGLLVKRHFHCARRNSKALFSQILLPAFFVCVAMTVALSVP EIGDLPL	1488
Db	432	QLKGWKLTTQQQFVALLWKRLLIARRSRKGFFAQIVLPAVFVCIALVFSLIVPPFGKYP	491
Qy	1489	VLSPSQYH-NYTQPRGNFIPIYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKS	1547
Db	492	ELQPWMYNEQYT-----FVSNDAPE-----DMGTQELLNALT KDPGFGTRCMEGN	536
Qy	1548	PANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPE	1607
Db	537	PIP-----DTPCL-----AGEE	548
Qy	1608	DLQAWNVS LPPTAGPEM-----WTSAPSLPRLVREPVRCTCSAQGTGFS---CPSSVGG-	1658
Db	549	D---WTISPVPQSI VDLFQNGNWTMKNPSP-----ACQCSSDKIKMLPVCPPGAGGL	598
Qy	1659	HPPQMRVVTGDILTDTIGHNVSEYLLFTSDRF-----RLHRYGAITFG-----	1701
Db	599	PPPQRKQKTADILQNL TGRNISDYLVKTYVQIIAKSLKNKIWNNEFRYGGFSLGVSNSQA	658
Qy	1702	-----NVLKSIPASFGTRAPPMVRKIA-----VRRAAQVFYNNKGYHSMPT	1742
Db	659	LPPSHEVNDAIKQMKKLLKLT KDT SADRFLSSLGREMAGLDTKNNVKVWFNNKGWHAISS	718
Qy	1743	YLNSLNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLS-LDYLLQGT DVVIAIFIIV	1801
Db	719	FLNVINAILRANLQKGE-NPSQYGITAFNHPLNLTKQQLSEVALMTTSVDVLVSICVIF	777
Qy	1802	AMSFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYLPATCCVILFV	1861
Db	778	AMSFVPASFVVFLIQERVSKAKHLQFISGVKPIYWLNSFVWDMCNVVPATLVIIIFIC	837
Qy	1862	FDLPAYTSPTNFPAVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITATV	1921
Db	838	FQQKSYVSSTNLPVLALLLLLYGWSITPLMYPASFVKIPSTAYVVLTSVNLFIGINGSV	897
Qy	1922	ATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMK	1981
Db	898	ATFVLELFTNNK-LNDINDILKSVFLIFPHFCLGRGLIDMVKNQAMADALERFGE-NRFV	955
Qy	1982	SPFEWDIVTRGLVAMAVEGVVGFLLT IMCQYNFLRRPQRMVSTKPVED-DVDVASERQR	2040
Db	956	SPLSWDLVGRNLFAMAVEGVVFFLITVLIQYRFFIRPRPVKAKLPPLNDEDEDVRRERQR	1015
Qy	2041	VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLG VNGAGKTSTFKM	2100
Db	1016	ILDGGGQNDILEIKELTKIYRRK---RKPAVDRICIGIPPGEFCFGLLG VNGAGKTSTFKM	1072
Qy	2101	LTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW	2160
Db	1073	LTGDTPVTRGDAFLNKNSILSNIHEVHQNMGYCPQFDAITELLTGREHVEFFALLRGVPE	1132
Qy	2161	KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK	2220

```

      |:  :  :||: || | || :| | |||||:|||| | :|||
Db      1133 KEVGKFGEWAIKRLGLVKYGEKYASNYSGGNKRKLSTAMALIGGPPVFLDEPTTGMDPK 1192

Qy      2221 ARRFLWNLILDIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG 2280
      ||||| | :| | |||||:||||| ||||:|||||
Db      1193 ARRFLWNCALSIVKEGRSVVLTSHSMEECEALCTRMAIMVNGRFRCLGSGVQHLKNRFGDG 1252

Qy      2281 YMITVR-TKSSQSVKDVVRFFNRNPFPEAMKERHHTKVQYQLKSEHISLAQVFSKMEQVS 2339
      | | | | | :  :| | | | | :|||:| | | | : |
Db      1253 YTIVVRIAGSNPDLKPVEFFGLAFPGSVLKEKHRNMLQYQLPSSLSSLARIFSILSQSK 1312

Qy      2340 GVLGIEDYSVSQTTLDNVFVNFAKKQSDN 2368
      | ||||| ||||| |||:
Db      1313 KRLHIEDYSVSQTTLDQVFNFAKDQSD 1341

```

RESULT 5

US-08-762-500-26

; Sequence 26, Application US/08762500

; Patent No. 6030806

; GENERAL INFORMATION:

; APPLICANT: Landes, Gregory M.

; APPLICANT: Burn, Timothy C.

; APPLICANT: Connors, Timothy D.

; APPLICANT: Dackowski, William R.

; APPLICANT: Van Raay, Terence J.

; APPLICANT: Klinger, Katherine W.

; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,

; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME

; NUMBER OF SEQUENCES: 83

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENZYME CORPORATION

; STREET: One Mountain Road

; CITY: Framingham

; STATE: Massachusetts

; COUNTRY: United States of America

; ZIP: 01701

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/762,500

; FILING DATE: 09-DEC-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/665,259

; FILING DATE: 17-JUN-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US96/10469

; FILING DATE: 17-JUN-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Dugan, Deborah A.

; REGISTRATION NUMBER: 37,315

; REFERENCE/DOCKET NUMBER: IG5-9.3

; TELECOMMUNICATION INFORMATION:



; TELEPHONE: (508) 872-8400  
 ; TELEFAX: (508) 872-5415  
 ; INFORMATION FOR SEQ ID NO: 26:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1375 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 US-08-762-500-26

Query Match 25.1%; Score 3173.5; DB 3; Length 1375;  
 Best Local Similarity 46.5%; Pred. No. 5.4e-274;  
 Matches 683; Conservative 205; Mismatches 372; Indels 209; Gaps 30;

Qy 980 MEEEPHTLPLVVCVDKLTQVYKDDKKLALNKLNLNLYENQVVSFLGHNGAGKTTTMSILT 1039  
 ||||| | | : | ||:| |:|:| |:| | | : | |||||  
 Db 2 MEEEPHTLRLGVSIQNLVKVYRDGMKVAVDGLALNFYEGQITSFLGHNGAGKTTTMSILT 61

Qy 1040 GLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSRLKSMAQEE 1099  
 |||||:| | | ||:| | |:|:| ||||| | |||:| |:| | : : :  
 Db 62 GLFPPTSGTAYILGKDIRSEMSSIRQNLGVCPQHNVLFDMLTVEEHIWFYARLKGLSEKH 121

Qy 1100 IRREMDKMIEDLEL-SNKRHSLVQTLSSGGMKRKLSVAIAFVGGSRAILDEPTAGVDPYA 1158  
 : : ||:| | : | : | | |||:| ||||:| ||||:| : |||||  
 Db 122 VKAEMEQMALDVGLPPSKLKSQTSQSLSSGMQRKLSVALAFVGGSKVVILDEPTAGVDPYS 181

Qy 1159 RRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPLFLKGTGTYGTYR 1218  
 || ||:| |:| | |||:| |||||:| ||||| | || ||| | ||  
 Db 182 RRGIWELLKYRQGRTIILSTHHMDEADILGDRIAIISHGKLCCVGSFLKNQLGTGY 241

Qy 1219 LTLVKRPAEPG-----GPQEPGLASSPPGRAPLSSCSELQVSQ 1256  
 ||||: | : || | | : |  
 Db 242 LTLVKKDVESLSSCRNSSSTVSLCKEDSVSQSSSDAGLSDHESDTLTIDVS--AISN 299

Qy 1257 FIRKHVASCLLVSDTSTELSYILPSEAAKGAFFERLFOHLERSLDALHLSSFGLMDTTLE 1316  
 ||||: || | ||:| || |||:| || || : | | :| |:| : |||  
 Db 300 LIRKHVSEARLVEDIGHELTYVLPYEAKEGAFVELFHEIDRLSDLGISSYGISETTLE 359

Qy 1317 EVFLKVSEEDQSLSENSEADVKESSRKDVLPGAEGPASGEGHAGNLARCELTQSQASLQSA 1376  
 |:| |:| | | | | : ||  
 Db 360 EIFLKVAEE-----SGVDA-ETSDGTLP----- 381

Qy 1377 SSVGSARGDEGAGYTDVYGDYRPLF-----DNPQDPD--NVSLQEVEAEALSRV-GQGSR 1428  
 || : | : || : | : || : : : | : || : | : ||  
 Db 382 -----ARRNRRA-----FGDKQSCSLHPFTEDDAVDPNDSIDPESRETDLLSGMDGKGSY 431

Qy 1429 KLDGGWLKVRQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPL 1488  
 : | | | : || || || |||:| | |:|:| || |:| : | || | | |  
 Db 432 QLKGWKLTQQQFVALLWKRLLIARRSRKGFFAQIVLPAVFVCIALVFSLVPPFPFGKYPSL 491

Qy 1489 VLSPSQYH-NYTQPRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKS 1547  
 | | |:| || : : : | | |:|:| | | |:| :  
 Db 492 ELQPMWYNEQYT-----FVSNDAP-----DMGTQELLNALT KDPPGFGTRCMEGN 536

Qy 1548 PANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDE 1607  
 | : || : : |

Db 537 PIP-----DTPCL-----AGEE 548

Qy 1608 DLQAWNVS LPPTAGPEM-----WTSAPSLPRLVREPVRCTCSAQGTGFS---CPSSVGG- 1658  
| | : | : : | | | | : | |

Db 549 D---WTISPVPQSI VDLFQNGNWTMKNPSP-----ACQCSSDKIKKMLPVCPPGAGGL 598

Qy 1659 HPPQMRVVTGDILT DITGHN VSEYLLFTSDRF-----RLHRYGAITFG----- 1701  
| | : | | : : | | : | : | | : |

Db 599 PPPQRKQKTADILQNL TGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNSQA 658

Qy 1702 -----NVLKSIPASFGTRAPPMVRKIA-----VRRAAQVFYNNKGYHSMPT 1742  
: | : : | | : : : : : : : : : : : : :

Db 659 LPPSHEVND AIKQMKLLKLT KDTSADRFLSSLGRFMAGLDTKNNVKVWFNNKGWHAISS 718

Qy 1743 YLNSLN NAILRANLPKSKGNPAAYGITVTNHPMNKTSASLS-LDYLLQGT DVVIAIFIIV 1801  
: | : | | | | | | | : | : | | | | | | : | : | : | : |

Db 719 FLNVIN NAILRANLQKGE-NPSQYGITAFNHPLNLTKQQLSEVALMTTSVDVLVSICVIF 777

Qy 1802 AMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYLPATCCVILFV 1861  
| | | | | | | | : | : | | | | : | | : | | | | | : |

Db 778 AMSFVPASFVFLIQERVS KAKHLQFISGVKPVIIYWLSNFWDMCNYVVPATLVIIIFIC 837

Qy 1862 FDLPAYTSPTNFP AVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITATV 1921  
| : | | | : | | | | | : | | | | : | | : | | | : |

Db 838 FQQKSYVSSTNLPVLALL LLLYGWSITPLMPASFVFKIPSTAYVVLTSVNLFINGSV 897

Qy 1922 ATFLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMK 1981  
| | : | : | : | : | | | | | : | | : | | : | : : | : : :

Db 898 ATFVLELFTNNK-LNDINDILKSVFLIFPHFCLGRGLIDMVKNQAMADALERFGE-NRFV 955

Qy 1982 SPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVPSTKPVED-DVDVASERQR 2040  
| | | : | | | | | | | : : | | | : : | : | | | |

Db 956 SPLSWDLVGRNLFAMAVEGVVFLITVLIQYRFFIRPRPVKAKLPPLNDEDEDVRRERQR 1015

Qy 2041 VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRCLGVRPGECFGLLG VNGAGKTSTFKM 2100  
: | | | : : : | : : : | | | : : | : | | | | : : |

Db 1016 ILDGGGQNDILEIKELTKIYRRK---RKA PAVDRICIGIPPGECFGLLG VNGAGKTSTFKM 1072

Qy 2101 LTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW 2160  
| | | | | : : : | : : : : | : : | | | : : | | :

Db 1073 LTGDTPTVTRGDAFLNKSILSNIHEVHQNMGYCPQFDAITELLTGREHVEFFALLRGVPE 1132

Qy 2161 KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK 2220  
| : : : | : | | : | | | | | : | | | : | | | | |

Db 1133 KEVGKFGEWAIRKLGLVKYGEKYASNYSGGNKRKLSTAMALIGPPVVLDEPTTGMDPK 1192

Qy 2221 ARRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG 2280  
| | | | | : : | | | | | : : | | | : | | | : | | |

Db 1193 ARRFLWNCALSIVKEGRSVVLTSHSMEECEALCTRLAIMVNGRFRCLGSVQHLKNRFGDG 1252

Qy 2281 YMITVR-TKSSQSVKDVVRFFNRNFP EAMLERHHTKVQYQLKSEHISLAQVFSKMEQVS 2339  
| | | | : : | | | | : : | | : | | | : |

Db 1253 YTIVVRIAGSNPDLKPQVEFFGLAFPGSVLKEKHRNMLQYQLPSSLSSILARIFSILSQSK 1312

Qy 2340 GVLGIEDYSVSQTTLDNVFNFAKKQSDN 2368  
| | | | | | | | | | | | |

Db 1313 KRLHIEDYSVSQTTL DQVFNFAKDQSD 1341

RESULT 6

US-08-665-259-25

; Sequence 25, Application US/08665259

; Patent No. 6028173

; GENERAL INFORMATION:

; APPLICANT: Landes, Gregory M.

; APPLICANT: Burn, Timothy C.

; APPLICANT: Connors, Timothy D.

; APPLICANT: Dackowski, William R.

; APPLICANT: Van Raay, Terence J.

; APPLICANT: Klinger, Katherine W.

; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,

; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME

; NUMBER OF SEQUENCES: 73

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENZYME CORPORATION

; STREET: One Mountain Road

; CITY: Framingham

; STATE: Massachusetts

; COUNTRY: United States of America

; ZIP: 01701

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/665,259

; FILING DATE: 17-JUN-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Dugan, Deborah A.

; REGISTRATION NUMBER: 37,315

; REFERENCE/DOCKET NUMBER: IG5-9.1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (508) 872-8400

; TELEFAX: (508) 872-5415

; INFORMATION FOR SEQ ID NO: 25:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1684 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-665-259-25

Query Match 20.7%; Score 2622; DB 3; Length 1684;

Best Local Similarity 34.0%; Pred. No. 2.1e-224;

Matches 638; Conservative 317; Mismatches 556; Indels 364; Gaps 45;

Qy 581 KGFPDEESIVNYTLNQAYQDNV--TVFASVIFQ---TRKDGSLPPHVHYKIR----- 627  
:| | | | : : | | : | | : | : |

Db 88 RGFPEKDFEDY----IRYDNCSSSVLAADVFEHPFNHSKEPLPLAVKYHLRFESYTRRNY 143

Qy 628 ---QNSSFTEK-----TNEIRRAYWRPG-----PNTGGRFYFLYGFWVIQDMMERAI 672  
| | | | : : | | | : | | | : | : | : |

Db 144 MWTQTGSFFLKETEGWHTTSLFPLFPNPGPRELTSPDGGEPGYIREGFLAVQHAVDRAIM 203

Qy 673 DTFVGHDVVEPGSY-----VQMFPPYPCYTRDDFLFVIEHMMPLCMVISWVYSVAMTIQH 726  
: | : : ||| : | | | : : | : : : | :

Db 204 EYHA--DAATRQLFQRLTVTIKRFYPFFIADPFLVAIQYQLPLLLLSFTYTALTARA 261

Qy 727 IVAEKEHRLKEVMKTMGLNNAVHWVAFITGFVQLSISVTALTAIL-----KYGQVLMHS 781  
: | | | | | : | | : : | | : | : : | : | |

Db 262 VVQEKERRLKEYMRMMGLSSWLHWSAWFLFLFLFLIAASFMTLLFCVKVKNVAVLSRS 321

Qy 782 HVVIIWFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHD 841  
: : | | : : | | | | : | | : | | | | : | | : | | : | | : :

Db 322 DPSLVLAFLLCFAISTISFSEFMVSTFFSKANMAAAFGGFLYFFTYIPYFFVAPR----YN 377

Qy 842 KITAFEKCIASLMSTTAFGLGSKYFALYEAVGVGIQWHTFSQSPVE-GDDFNLLLAVTML 900  
: | : | : | : | : : : | | : | | | | | | | | : | |

Db 378 WMTLSQKLCSCLLSNVAMAMGAQLIGKFEAKGMIQWRDL-LSPVNVDDDFCFGQVLGML 436

Qy 901 MVDVVYVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSV 960  
: : | : | : | : | : | | | : | : | : | : | : | : | : | :

Db 437 LLDSVLYGLVTWYMEAVFPQGFQVPQWYFFIMPSYWCCKPRAVAGK----- 483

Qy 961 MEEDQACAMESRRFEETRGMEEEPTHLPLVVCVDKLTKVYK--DDKKLALNKLNLNLYEN 1018  
| | : : : | | | | | : : | : | : : : : | : | : | | |

Db 484 -EEEDSDPEKALRNEY---FEAEPEDIVAGIKIKHLSKVFRVGNKDRAVRDLNLNLYEG 539

Qy 1019 QVVSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFD 1078  
| : | | | | | | | : | : | | | | | | | | : | : | : | : | : | : | :

Db 540 QITVLLGHNGAGKTTTSLMLTGLFPPTSGRAYISGYEISQDMVQIRKSLGLCPQHDILFD 599

Qy 1079 RLTVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAF 1138  
| | | | | : | : | : : : : | : : : | : : | : | : | : | : | :

Db 600 NLTVAEHLFYAQLKGLSRQKCPPEEVKQMLHIIGLEDKWNRSRFLSSGGMRRKLSIGIAL 659

Qy 1139 VGGSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHG 1198  
: | : : | | | : | : | : | : : | | : | : | | | | | : : |

Db 660 IAGSKVLILDEPTSGMDAISRRAIWDLQKQSDRTIVLTTHFMDEADLLGDRIAIMAKG 719

Qy 1199 KLKCCGSPLFLKGTYG DYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFI 1258  
: | : | | | | | | | | : | : | : | : | : | : | : | :

Db 720 ELQCCGSSLFLKQKYGAGYHMTLVKEP-----HCNPEDISQLV 757

Qy 1259 RKHVASCLLVSDTSTELSYILPSEAAKKGA FERLFQHLERSLDALHLSSFGLMDTTLEEV 1318  
| | : | | | | : | : : | | | | : | : | : | | : | : | |

Db 758 HHHVPNATLESSAGAELS FILPRESTHR--FEGLEFAKLEKKQKELGASFGASITTEEV 815

Qy 1319 FLKVSEEDQSLENSEADV KESRKDVLPGAEGPASGEGHAGNLARCELTQSQASLQSASS 1378  
| : | : | : | : | : : | : | : | : | : | : | : | :

Db 816 FLRVGK----LVDSSMDIQAIQ---LPALQ--YQHERRASDWAVDSNL----- 854

Qy 1379 VGSARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALS RVGQGSRKLDGGW-LKV 1437  
| : : | | : | | | | : | : | : | | : | |

Db 855 CGAMDPSDGIG-----ALIEEERTAV-----KLNTGLALHC 885

Qy 1438 RQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYHN 1497  
: | | : : | : | : | : | : | : | : | : | : | : | :

Db 886 QQFWAMFLKKAAYS WREWKMVAAQVLVPLTCVTLLALLAINYSSELFD DDPMLRLTLGEY-- 943

Qy	1498	YTQPRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCV-LKSPANGSLGPT	1556
Db	944	-----GRTVVPFVPGTSQLGQQ	961
Qy	1557	LNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDELQAWNVS	1616
		:	
Db	962	LS-----	963
Qy	1617	PPTAGPEMWTSAAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTDITG	1676
		::   :   :     :	
Db	964	-----EHLKDALQAEG-----QEPREVLGDL-----	984
Qy	1677	HNVSLEYLLFTSDRFLHRYGAITFG--NVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNN	1734
		:   :   : :     :       :   :   :	
Db	985	---EEFLIFRA-----SVEGGGFNERCLVAASF-----RDVGERTVVNALFNN	1024
Qy	1735	KGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQG-TDV	1793
		:         : :   :         : :   :   :	
Db	1025	QAYHSPATALAVVDNLLF-----KLLCGPHA-SIVVSNFQPRQPRALQAQKQFNEGRKGF	1078
Qy	1794	VIAIFIIIVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYLVPAT	1853
		: :     : :   :   :   :   :   :   :   :   :	
Db	1079	DIALNLLFAMAFLASTFSILAVSERAVQAKHVQFVSGVHVASFWSALLWDLISFLIPSL	1138
Qy	1854	CCVIILFVFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINL	1913
		: : :   :   :   :         :   :   :   :   :	
Db	1139	LLLTVFKAFDVRAFTRDGHMADTLLLLLLLYGWAIIPLMYLMNFFFLGAATAYTRLTIFNI	1198
Qy	1914	FIGITATVATFLLQLFEHDKDLKV--VNSYLKSCFLIFPNYNLGHGLMEMAYNEY-----	1966
		: :   :   :   :   :   :   :   :	
Db	1199	LSGI---ATFLMVTIMRIPAVKLEELSKTLDHVFVLVLPNHCLGMAVSSF-YENYETRRY	1253
Qy	1967	-----INEYYAKIGQFDMKSPFEWDI--VTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQ	2019
		: :   : :   :   :   :   :   :   :   :	
Db	1254	CTSSEVAAHYCKKYNIQYQENFYAWSAPGVGRFVASMAASGCAYLILLFLIETNLLQRLR	1313
Qy	2020	-----RMPVSTKPVEDDQVAVSERQVRVLRGDADNDM---VKIENLTKV	2059
		: :         :   : :   :   :	
Db	1314	GILCALRRRRTLTELYTRMPV----LPEDQDVADERTRILAPSPDSSLHTPLIIEKELSKV	1369
Qy	2060	YKSRKIGRILAVDRLCLGVRPGECFGLLGVNGAGKTSTFKMLTGDESTTGGEAFVNGHSV	2119
		:   :           :           :           :         :	
Db	1370	YEQRV--PLLAVDRLSLAVQKGEFCFGLLGFGAGKTTTFKMLTGEESLTSGDAFVGGHRI	1427
Qy	2120	LKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKY	2179
		: : :   :           :       :       :   :   :   :	
Db	1428	SSDVGVKVRQIRIGYCPQFDALLDHMTGREMLVMYARLRGIPERHIGACVENTLRGLLLEPH	1487
Qy	2180	ADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIKTGRSV	2239
		:                           :       : : : : :	
Db	1488	ANKLVRTYSGGNKRKLSTGIALIGEPVIFLDEPSTGMDPVARRLLWDTVARARESGKAI	1547
Qy	2240	VLTSHSMEECEALCTRLAIMVNGRLRLCLGSIQHLKNRFGDGYMITVRTKS---SQSVKDV	2296
		: :                       : :       : :   : : :	
Db	1548	IITSHSMEECEALCTRLAIMVQGFQKCLGSPQHLKSKFGSGYSLRAKVQSEGQOEALIEEF	1607

Qy 2297 VRFFNRNFP EAM LKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDN 2356  
 | : || ::| | | :| |::| :| :| ::| | | :| :  
 Db 1608 KAFVDLT FPGSVLEDEHQGMVHYHLPGRDLSWAKVFGILEKAKEKYGVDDYSVSQISLEQ 1667  
 Qy 2357 V FVNFAKKQSDNLEQ 2371  
 ||::|| | | :  
 Db 1668 VFLSFAHLQPPTAEE 1682

RESULT 7

US-08-762-500-25

; Sequence 25, Application US/08762500

; Patent No. 6030806

; GENERAL INFORMATION:

; APPLICANT: Landes, Gregory M.

; APPLICANT: Burn, Timothy C.

; APPLICANT: Connors, Timothy D.

; APPLICANT: Dackowski, William R.

; APPLICANT: Van Raay, Terence J.

; APPLICANT: Klinger, Katherine W.

; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,

; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME

; NUMBER OF SEQUENCES: 83

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENZYME CORPORATION

; STREET: One Mountain Road

; CITY: Framingham

; STATE: Massachusetts

; COUNTRY: United States of America

; ZIP: 01701

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/762,500

; FILING DATE: 09-DEC-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/665,259

; FILING DATE: 17-JUN-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US96/10469

; FILING DATE: 17-JUN-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Dugan, Deborah A.

; REGISTRATION NUMBER: 37,315

; REFERENCE/DOCKET NUMBER: IG5-9.3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (508) 872-8400

; TELEFAX: (508) 872-5415

; INFORMATION FOR SEQ ID NO: 25:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1684 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein  
US-08-762-500-25

Query Match 20.7%; Score 2622; DB 3; Length 1684;  
Best Local Similarity 34.0%; Pred. No. 2.1e-224;  
Matches 638; Conservative 317; Mismatches 556; Indels 364; Gaps 45;

```
Qy      581 KGFPDEESIVNYTLNQAYQDNV--TVFASVIFQ---TRKDGSLPPHVHYKIR----- 627
      :||| |:  :|      ||  :| |:|:|:      ||  | | :|
Db      88 RGFPSEKDFEDY----IRYDNCSSSVLAAVVFEHPPFNHSKEPLPLAVKYHLRFSYTRRNY 143

Qy      628 ---QNSSFTEK-----TNEIRRAYWRPG-----PNTGGRFYFLYGFWIQDMMERAI 672
      |  | |  |      | :  : ||      | : |  | | : | :| | | :
Db      144 MWTQTGSFFLKETEGWHTTSLFPLFPNPGPRELTSPDGGEPGYIREGFLAVQHAVDRAIM 203

Qy      673 DTFVGHDVVEPGSY-----VQMFYPYCYTRDDEFLFVIEHMMPLCMVISWVYSVAMTIQH 726
      :  |      :      :| | | :  | | | | :| :| :| :| :| :| :
Db      204 EYHA--DAATRQLFQRLTVTIKRFPPFPFIADPFLVAIQYQLPLLLLLLSFTYTALTIA 261

Qy      727 IVAEKEHRLKEVMKTMGLNNAVHWVAFITGFVQLSISVTALTAIL-----KYGQVLMHS 781
      :| | | | | | | :| | | :| | | :| :| :| :| :| :| :| :| :| :| :|
Db      262 VVQEKERRLKEYMRMMGLSSWLHWSAWFLLFFLFLLLIAASEFMTLLFCVKVKPNVAVLSRS 321

Qy      782 HVVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHD 841
      :| | | :| :| | | | | :| | | :| | | :| | | :| | | :| | | :|
Db      322 DPSLVLAFLLCFAISTISFSFMVSTFFSKANMAAAFGGFLYFFTYIPYFFVAPR----YN 377

Qy      842 KITAFEKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVE-GDDFNLLLAVTML 900
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      378 WMTLSQKLCSCLLSNVAMAMGAQLIGKFEAKGMGIQWRDL-LSPVNVDDDFCFGQVLGML 436

Qy      901 MVDVVYGYLTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSV 960
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      437 LLDSVLYGLVTWYMEAVFPGQFQVPQWPYFFIMPSYWCCKPRAVAGK----- 483

Qy      961 MEEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTQVYK--DDKKLALNKLNLNLYEN 1018
      || :  :  :| |  | | | :  :| :| :| :| :| :| :| :| :| :| :|
Db      484 -EEEDSDPEKALRNEY---FEAEPEDLVAGIKIKHLSKVFRVGNKDRAAVRDLNLNLYEG 539

Qy      1019 QVVSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFD 1078
      | :  | | | | | | | | :| :| | | | | | | | | :| :| :| :| :| :| :|
Db      540 QITVLLGHNGAGKTTTSLMLTGLFPPTSGRAYISGYEISQDMVQIRKSLGLCPQHDILFD 599

Qy      1079 RLTVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAF 1138
      || | | | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      600 NLTVAEHLFYAQLKGLSRQKCPPEEVQMLHIIGLEDKWNRSRFLSGGMRRKLSIGIAL 659

Qy      1139 VGSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHMHDEADLLGDRIAIISHG 1198
      : | | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      660 IAGSKVLILDEPTSGMDAISRAIWDLLQRQKSDRTIVLTTHFMDEADLLGDRIAIMAKG 719

Qy      1199 KLKCCGSPLFLKGTGTYGDRYRLTVKRPAPGPGQEPGLASSPPGRAPLSSCSELQVSQFI 1258
      :| :| | | | | | | | | :| :| :| :| :| :| :| :| :| :| :|
Db      720 ELQCCGSSLFLKQKYGAGYHMTLVKEP-----HCNPEDISQLV 757

Qy      1259 RKHVASCLLVSDTSTELSYILPSEAAKKGAFERLFQHLERSLDALHLSSFGLMDTTLEEV 1318
      || :  | |  | | :| :| :| :| :| :| :| :| :| :| :| :| :|
```

Db 758 HHHVPNATLESSAGAELSFILPRESTR--FEGLFAKLEKKQKELGASFSGASITTMEEV 815  
 Qy 1319 FLKVSEEDQSLENSEADVKESRKDVLPGAEGPASGEGHAGNLARCSELTQSQASLQSASS 1378  
 ||:| : | :| |:: : || : | | : | | |  
 Db 816 FLRVGK----LVDSSMDIQAIQ---LPALQ--YQHERRASDWAVDNL----- 854  
 Qy 1379 VGSARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVQGSRKLDGGW-LKV 1437  
 |: :| | | :| | | |: ||: | |  
 Db 855 CGAMDPSDGIG-----ALIEEERTAV-----KLNTGLALHC 885  
 Qy 1438 RQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVP EIGDLPPIVLSPSQYHN 1497  
 :|| : :| : | : :|:| | :| : | | | | : :|  
 Db 886 QQFWAMFLKKAAYSWREWKMVAAQVLVPLTCVTLALLAINYSSELFDDPMLRLTLGEY-- 943  
 Qy 1498 YTQPRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCV-LKSPANGSLGPT 1556  
 | | | | | |  
 Db 944 -----GRTVVPFSPVPGTSQLGQQ 961  
 Qy 1557 LNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDQLQAWNVSL 1616  
 |:  
 Db 962 LS----- 963  
 Qy 1617 PPTAGPEMWTSAFSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTDTIG 1676  
 | : : | :| : | | ||:  
 Db 964 -----EHLKDALQAEG-----QEPREVLGDL----- 984  
 Qy 1677 HNVSEYLLFTSDRFRHLHRYGAITFG--NVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNN 1734  
 |:|:| : : : | | : || | : | :||  
 Db 985 ---EEFLIFRA-----SVEGGGFNERCLVAASF-----RDVGERTVVNALFNN 1024  
 Qy 1735 KGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQG-TDV 1793  
 : ||| | | : :| : | | | | | :| : : : | : |  
 Db 1025 QAYHSPATALAVVDNLIF-----KLLCGPHA-SIVVSNFPQPRALQAAKDQFNEGRKGF 1078  
 Qy 1794 VIAFIIVAMSEVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPAT 1853  
 ||: : ||:|: :| : |:|: : ||:| ||| : :||: :||: :|:|:  
 Db 1079 DIALNLLFAMAFLASTFSILAVSERAVQAKHVQFVSGVHVASFWSALLWDLISFLIPSL 1138  
 Qy 1854 CCVILFVFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWFVPSAYVFLIVINL 1913  
 : : ||: |:| : | | |||||:| |:| :|: :|| | : |:  
 Db 1139 LLLVVFKAFDVRAFTRDGHMADTLLLLLLLYGWAIIPLMYLMNFFFLGAATAYTRLTIFNI 1198  
 Qy 1914 FIGITATVATFLLQLFEHDKDLKV--VNSYLKSCFLIFPNYNLGHGLMEMAYNEY----- 1966  
 || |||: :|: : : | ||: ||: | : | |  
 Db 1199 LSGI----ATFLMVTIMRIPAVKLEELSKTLDHVFLVLPNHCLGMAVSSF-YENYETRRY 1253  
 Qy 1967 -----INEYYAKIGQFDKMKSPFEWDI--VTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQ 2019  
 : : | | : : | | | : || | : | : | | : |  
 Db 1254 CTSSEVAHYCKKYNIQYQENFYAWSAPGVGRFVASMAASGCAYLILLFLIETNLLQRLR 1313  
 Qy 2020 -----RMPVSTKPVEDDQDVASERQVLRGDADNDM---VKIENLTKV 2059  
 |||| : :| ||| || |:| | : : : | : |:  
 Db 1314 GILCALRRRRTLTelyTRMPV----LPEDQDVADERTRILAPSPDSSLHTPLIIKELSKV 1369  
 Qy 2060 YKSRKIGRILAVDRCLGVRPGECFGLLGVNGAGKTSTFKMLTGDESTTGGEAFVNGHSV 2119  
 |: | : ||||| | | : ||||| |||||:|||||:| | | :||| | :  
 Db 1370 YEQRV--PLAVDRLSLAVQKGEFCFGLLGFGAGKTTTFKMLTGEESLTSGDAFVGGHRI 1427



Qy 2120 LKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKY 2179  
 :: :||:| :||||| ||| | :| || | :| ||||| : | : | | :  
 Db 1428 SSDVGKVRQRIGYCPQFDALLDHMTGREMLVMYARLRGIPERHIGACVENTLRGLLLEPH 1487  
 Qy 2180 ADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIKTGRSV 2239  
 |:| ||||| ||||| ||||| || |||||:||||| ||| ||: : :|:::  
 Db 1488 ANKLVRTYSGGNKRKLSTGIALIGEPAVIFLDEPSTGMDPVARRLWDTVARARESGKAI 1547  
 Qy 2240 VLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVRTKS---SQSVKDV 2296  
 ::||| ||||| ||||| ||| :||| ||||:|| || : : :| ::::  
 Db 1548 IITSHSMEECEALCTRLAIMVQGFQKCLGSPQHLKSKFGSGYSLRAKVQSEGQQEALEEF 1607  
 Qy 2297 VRFFNRNFPPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDN 2356  
 | : || :||:| :| | | :| :|| :| :| :||| || :| :  
 Db 1608 KAFVDLTFPGSVLEDEHQGMVHYHLPGRDLSWAKVFGILEKAKEKYGVDDYSVSQISLEQ 1667  
 Qy 2357 VFNFAKKQSDNLEQ 2371  
 ||:|| | | :  
 Db 1668 VFLSFAHLQPPTAEE 1682

RESULT 8

US-08-762-500-75

; Sequence 75, Application US/08762500

; Patent No. 6030806

; GENERAL INFORMATION:

; APPLICANT: Landes, Gregory M.

; APPLICANT: Burn, Timothy C.

; APPLICANT: Connors, Timothy D.

; APPLICANT: Dackowski, William R.

; APPLICANT: Van Raay, Terence J.

; APPLICANT: Klinger, Katherine W.

; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,

; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME

; NUMBER OF SEQUENCES: 83

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENZYME CORPORATION

; STREET: One Mountain Road

; CITY: Framingham

; STATE: Massachusetts

; COUNTRY: United States of America

; ZIP: 01701

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/762,500

; FILING DATE: 09-DEC-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/665,259

; FILING DATE: 17-JUN-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US96/10469

```
; FILING DATE: 17-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IG5-9.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:-
; LENGTH: 1704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-762-500-75
```

```
Query Match          20.7%; Score 2622; DB 3; Length 1704;
Best Local Similarity 34.0%; Pred. No. 2.2e-224;
Matches 638; Conservative 317; Mismatches 556; Indels 364; Gaps 45;
```

```
Qy      581 KGFPDEESIVNYTLNQAYQDNV--TVFASVIFQ---TRKDGS LPPHVHYKIR----- 627
      :||| |:  :|      ||  :| |:|:|:      ||  ||  :|
Db      108 RGFPEKDFEDY----IRYDNCSSSVLAADVFEHPFNHSKEPLPLAVKYHLRF SYTRRNY 163

Qy      628 ---QNSSFTEK-----TNEIRRAYWRPG-----PNTGGRFYFLYGFVWIQDMMERAI 672
      |  ||  |      |  :  :  ||      | :  |  |  ||: :|  :|||:
Db      164 MWTQTGSFFLKETEGWHTTSLFPLFPNPGPRELTSPDGGEPGYIREGFLAVQHAVDRAIM 223

Qy      673 DTFVGHDVVEPGSY-----VQMFPPYPCYTRDDFLFVIEHMMPLCMVISWVYSVAMTIQH 726
      :  |      :      : : ||| :  ||  | : : ||  : :| :| :  :
Db      224 EYHA--DAATRQLFQRLTVTIKRFPPYPPFIADPFLVAIQYQLPLLLLSFTYTALTIARA 281

Qy      727 IVAEKEHRLKEVMKTMGLNNAVHWVAWFITGFVQLSISVTALTAL-----KYGQVLMHS 781
      :| ||| ||| |: |||: :|| |||:  | : | : : :| :      ||  |
Db      282 VVQEKERRLKEYMRMMGLSSWLHWSAWFLFLFLLLIAASFMTLLFCVKVKNVAVLSRS 341

Qy      782 HVVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEV AHD 841
      : :  ||  :|: :|| | | :||  :||| :|:| ||  :||  :|:| :||  : :
Db      342 DPSVLVLAFLLCFAISTISFSEFMVSTFFSKANMAAFGGFLYFFTYIPYFFVAPR----YN 397

Qy      842 KITAFEKCIASIMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVE-GDDFNLLLA VTML 900
      :|  :|  : |:| |  :|: :|  | :|||  |||  |||  : ||
Db      398 WMTLSQKLCSCLLSNVAMAMGAQLIGKFEAKGMGIQWRDL-LSPVNVD DFCFGQVLGML 456

Qy      901 MVDVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSV 960
      : :| :|: ||: ||| :|| ||  :|: ||| :  ||| |  |  | :
Db      457 LLDSVLYGLVTWYMEAVFPGQFGVPQWPYFFIMPSYWCGKPRAVAGK----- 503

Qy      961 MEEDQACAMESRRFEETRGMEEEPTHLPLVVCVDKLTKVYK--DDKKLALNKL SINLYEN 1018
      ||: :  : : | |  | ||  |  : :  | :||: :  : | : | :|||
Db      504 -EEEDSDPEKALRNEY---FEAEPEDLVAGIKIKHLSKVFRVGNKDRAAVRDLN LNLYEG 559

Qy      1019 QVVSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQH NVLFD 1078
      | :  ||| ||| ||| :|: ||| ||| | | | :| :| :||: ||: ||| :|||
Db      560 QITVLLGHNGAGKTTTSLMLTGLFPPTSGRAYISGYEISQDMVQIRKSLGLCPQH DILFD 619

Qy      1079 RLTVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAF 1138
```

Db 620 NLTVAEHLFYFAQLKGLSRQKCPEEVKQMLHIIGLEDKWNRSRFLSGGMRRKLSIGIAL 679  
 Qy 1139 VGGSRAILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHG 1198  
 : ||: :|||:|:|:| :|||:|:|:| :| :||:|:| :|||:|:|:|:|:|:|  
 Db 680 IAGSKVLILDEPTSGMDAISRRAIWDLQRQKSDRTIVLTTHFMDEADLLGDRIAIMAKG 739  
 Qy 1199 KLKCCGSPLFLKGTYG DG YRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFI 1258  
 :|:|||| ||| || || :||| | | : :|| :  
 Db 740 ELQCCGSSLFLKQKYGAGYHMTLVKEP-----HCNPEDISQLV 777  
 Qy 1259 RKHVASCLLVSDTSTELSYILPSEAAKKGA FERLFQH LERSLDALHLSSFGLMDTTLEEV 1318  
 || : | | |||:| | : : || | | : | :|| | | :|||  
 Db 778 HHHVPNATLESSAGAELSFILPRESTR--FEGLEFAKLEKKQKELG IASFGASITMEEV 835  
 Qy 1319 FLKVSEEDQSLENSEADV KESRKDVLPGAEGPASGEGHAGNLARCSELTQSQASLQSASS 1378  
 ||:| : | :| |:: : || : | | : | | |  
 Db 836 FLRVGK---LVDSSMDIQAIQ---LPALQ--YQHERRASDWAVDSNL----- 874  
 Qy 1379 VGSARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQVEAEALSRVGQGSRKLDGGW-LKV 1437  
 | : : | | : | | | | : | | : | | |  
 Db 875 CGAMDPDSDGIG-----ALIEEERTAV-----KLNTGLALHC 905  
 Qy 1438 RQFHGLLVKRFHFCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYHN 1497  
 :|| : :| : | : :|:| | : : | : | | | : :|  
 Db 906 QQFWAMFLKKAAYSWREWKMVAAQVLVPLTCVTLALLAINYSSELFDDPMLRLTLGEY-- 963  
 Qy 1498 YTQPRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCV-LKSPANGSLGPT 1556  
 | | | | | |  
 Db 964 -----GRTVVPFSVPGTSQLGQQ 981  
 Qy 1557 LNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPEDLDQAWNVS L 1616  
 | :  
 Db 982 LS----- 983  
 Qy 1617 PPTAGPEMWTSA PSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTDITG 1676  
 | : : | :| : | | ||:  
 Db 984 -----EHLKDALQAEG-----QEPREVLGDL----- 1004  
 Qy 1677 HNVSEYLLFTSDRFR LHRYGAITFG--NVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNN 1734  
 |:|:| : : : | | : || | : | :||  
 Db 1005 ---EEFLIFRA-----SVEGGGFNERCLVAASF-----RDVGERTVVNALFNN 1044  
 Qy 1735 KGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQG-TDV 1793  
 : ||| | | :| : | | | | | : : | :|  
 Db 1045 QAYHSPATALAVVDNLLF-----KLLCGPHA-SIVVS NFPPQPR SALQA AKDQFNEGRKGF 1098  
 Qy 1794 VIAIFIIVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPAT 1853  
 ||: : ||:| : :| : |:|: :|||:| | | : :||: :|:|:|:|:|:  
 Db 1099 DIALNLLFAMAFLASTFSILAVSERAVQAKHVQFVSGVHVASFWSALLWDLISFLIPSL 1158  
 Qy 1854 CCVIILFVFDLPAYTSPTNFFAVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINL 1913  
 : : ||: | :| : | | |||:| | :| :| :|| | : |:  
 Db 1159 LLLVVFKA F DVRAFTRDGHMADTLLLLLLLYGWAIPLMYLMNFFFLGAATAYTRLTIFNI 1218  
 Qy 1914 FIGITATVATFLLQLFEHDKDLKV--VNSYLKSCFLIFPNYNLGHGLMEMAYNEY----- 1966  
 || |||: :| : : | ||: ||: || : | |

Db 1219 LSGI----ATFLMVTIMRIPAVKLEELSKTLDHVFLVLPNHCLGMAVSSF-YENYETRRY 1273

Qy 1967 -----INEYYAKIGQFDKMKSPFEWDI--VTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQ 2019  
: :| | : : : | | : :|| | :| : : | | :| :

Db 1274 CTSSEVAAHYCKKYNIQYQENFYAWSAPGVGRFVASMAASGCAYLILLFLIETNLLQRLR 1333

Qy 2020 -----RMPVSTKPVEDDQDVASERQVRVLRGDADNDM---VKIENLTKV 2059  
|||| : :| ||| || | :| | : : : | :| :||

Db 1334 GILCALRRRRTLTELYTRMPV----LPEDQDVADERTRILAPSPDSLLHTPLIIKELSKV 1389

Qy 2060 YKSRKIGRILAVDRCLCLGVRPGECFLLGVNGAGKTSTFKMLTGDESTTGGEAFVNGHSV 2119  
| : | :| |||| | :| ||||| ||||| :||| | :||| || :

Db 1390 YEQRV--PLAVDRSLAVQKGECFLLGFNGAGKTTTFKMLTGEESLTSGDAFVGGHRI 1447

Qy 2120 LKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKY 2179  
: :| :| :| |||| ||| :| ||| :| |||| : | :| | | :

Db 1448 SSDVGKVRQRIGYCPQFDALLDHMTGREMLVMYARLRGI PERHIGACVENTLRGLLLEPH 1507

Qy 2180 ADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARFLWNLILDLIKTGRSV 2239  
| :| ||||| ||||| |||| | ||||| :||| ||| || : : : :| :| :| :

Db 1508 ANKLVRTYSGGNKRKLSTGIALIGEPAVIFLDEPSTGMDPVARRLLWDTVARARESGKAI 1567

Qy 2240 VLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVRTKS---SQSVKDV 2296  
: :||| ||||| ||||| ||| :| ||| :| ||| :| :| :| :| :| :| :

Db 1568 IITSHSMEECEALCTRLAIMVQGFQKCLGSPQHLKSKFGSGYSLRAKVQSEGQQEALIEEF 1627

Qy 2297 VRFFNRNFPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDN 2356  
| : || :| :| | | | :| | :| :| :| :| :| :| :| :

Db 1628 KAFVDLTFPGSVLEDEHQGMVHYHLPGRDLWSAKVFGILEKAKEKYGVDDYSVSQISLEQ 1687

Qy 2357 VFNFAKKQSDNLEQ 2371  
|| :|| | | :

Db 1688 VFLSFAHLQPPTAEE 1702

# RESULT 9

US-09-328-352-7592

; Sequence 7592, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 7592

; LENGTH: 589

; TYPE: PRT

; ORGANISM: Acinetobacter baumannii

US-09-328-352-7592

Query Match 2.9%; Score 363.5; DB 4; Length 589;

Best Local Similarity 31.4%; Pred. No. 6.7e-23;

Matches 102; Conservative 59; Mismatches 131; Indels 33; Gaps 7;



Db 74 GTCLGMDIFTQREKIKKKIGYMTQYFSMWGNLTIRENLLFIARLYSL--DRRRERVERAL 131

Qy 1109 EDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVGGSRAIILDEPTAGVDPYARRAIWDLILK 1168  
 :| |: :| | : ||| |:::| : : ||||| ||| | : :  
 Db 132 SELGLTARQHQLAKELSGGWKQRMALAACMLHEPVLLFLDEPTAGVDPKARREFWQMLHQ 191

Qy 1169 YKP-GRTILLSTHHMDEADLLGDRIAIISHGKLCCKGS-----PLFLKGTYGDSYRL 1219  
 | ::|:|:|:|:|: : :| :|:|:| | : | | | |  
 Db 192 LSDRGISLLVSTHYMDEAERC-HKVAYLSYGRLLANGTIASIIASQNLITMRTSGAG--L 248

Qy 1220 TLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHAVASCLLVSDTSTELSYIL 1279  
 ||:: :|:| | : | :|  
 Db 249 TLLE-----SQLQRLPDIEQTVI-----FGNQLYITS 275

Qy 1280 PSEAAKKGAFERLFQHLERSLDALHLSFGLMDTTLEEVF 1319  
 || | | || :: : | :|| ||: |  
 Db 276 RDEAKLKSA---LFAFTQQGYE-----FCKVDNLEDAF 306

# RESULT 11

US-09-252-991A-31957

; Sequence 31957, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 31957

; LENGTH: 345

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-31957

Query Match 2.8%; Score 359.5; DB 4; Length 345;

Best Local Similarity 27.0%; Pred. No. 5.4e-23;

Matches 87; Conservative 78; Mismatches 134; Indels 23; Gaps 7;

Qy 990 VVCVDKLTQVYKDDKKLALNKLNLNLYENQVVSFLGHNGAGKTTTMSILTGLFPPTSGSA 1049  
 :: :|:|:| : : :| || : :| || |||||:|: :|:| |::||  
 Db 38 MIDIDRLSKRFSG--RTVVNDLSFRIDRGEIVGLLGPNAGACKSTTLKMLSGFLAPSAGSV 95

Qy 1050 TIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSRLKSMAQEEIRREMDKMIE 1109  
 |:| |: : : :| :| |: : :|:| | | : : : | || :| : :  
 Db 96 RIFGFDMQDKARQAQKLIGYLPENAPSYGEMTVEGFLAFVASIRDYSGREKRRRIDSAMD 155

Qy 1110 DLELSNKRHSLVQTLSSGGMKRKLSVAIAFVGGSRAIILDEPTAGVDPYARRAIWDLILKY 1169  
 :|| ::| |:::| | ||:::| | : :||| |:| : : |:  
 Db 156 CMELRDERRSIIETLSKGFKRRVALAQAILHDPELLLLDEPTDGLDPNQKHQVRQLVKNL 215

```

Qy      1170 KPGRITILLSTHHMDEADLLGDRIAIISHGKCLKCCGSP--LFLKGTYGDDGYRLTLVKRPAE 1227
      : ::||| ::| : | :|: |:| :| | : | |: :: | :
Db      216 SESKIVVISTHILEEVSMCSRALVINGGRLLADNTPGELRTRSRYYHHAVALS-IEAPVD 274

Qy      1228 PGG----PQEPGLASSPPGRAPLSSCSE--LQVSQFIRKHV-ASCLLVSDTSTELSYILP 1280
      | | | : | | : : :|: : : : | || ||
Db      275 PLAIAMLPGVAGIEGRPDRAGTLTILARPGVQILPALNRLIHGSGWRVSGVRTE----- 328

Qy      1281 SEAAKKGAFERLFQHLERSLDA 1302
      | | :|: | | |
Db      329 -----HGQLEEVFROLTRETTPA 345

```

US-09-543-681A-4646  
; Sequence 4646, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS  
MIRABILIS FOR  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 4646  
; LENGTH: 532  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-4646

Qy	1984	FEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDVDVA-----	2035
Db	223	FDW-----LVAMDAGKV-----LATGHAEELKAQTATDELEAAFIELLPEE	263
Qy	2036	--SERQRVL---RGDADNDMVKIE--NLTKVYKSRKIGRILAVDRICLGVRPGECFGLLG	2088
Db	264	KRKNHQKVIIPPRDKSDDDDIIAIEAKELT-----MRFGQFVAVDHVSFRIPKGEIFGFLG	318
Qy	2089	VNGAGKTSTFKMLTGDESTTGGAEFVNGHSLKELLQVQQSLGYCPQCDALFDELTAREH	2148
Db	319	SNGCGKSTTMKMLTGLLEASEGRAWLFQGVEVDPKDIETRKRVGYSQAFSLYSELTVRQN	378
Qy	2149	LQLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFI	2208
Db	379	LELHAKLFHIPEQDIPQRIKEMSERFNLTDEDMPDGLPLGIRQLSLAVAVIHKPEML	438
Qy	2209	FLDEPTTGMDPKARRFLWNLILDLI-KTGRSVLTSMSMEECEALCTRLAIMVNGRL	2264

## RESULT 13

US-09-252-991A-28171

; Sequence 28171, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 28171  
 ; LENGTH: 788  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-28171

Query Match 2.7%; Score 341.5; DB 4; Length 788;  
 Best Local Similarity 33.6%; Pred. No. 1.1e-20;  
 Matches 95; Conservative 45; Mismatches 110; Indels 33; Gaps 6;

Qy 2016 RRPQ--RMPVSTKPVEDDDVA-----SERQVLRGDADN-----DMVKI 2053  
 ||| :| | || | :| | ||| :|:|  
 Db 431 RRPADLAVPAGTPAVEQPEHRAAALGPALPGGALARRARQPSADADARRGAEAVSLVEI 490  
 Qy 2054 ENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLGVNGAGKTSTFKMLTGDESTTGGEAF 2113  
 : | | : | : | | : ||| |||| ||||| :| | : : |  
 Db 491 DGATLRY-----GALTALSGLDLRLPEGEVLGLLGHNGAGKTTTIKLVGLLAPSEGRVR 545  
 Qy 2114 VNGHSLVKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEK 2173  
 | || : : : ||| | : : : | | : : ||| : : | : | :  
 Db 546 VLGHDA--RSLEARRQLGYLPENVTFYPLSGAETLRHFARLKGVAPEAAARL----LEQ 599  
 Qy 2174 LELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARREFLWNLILDLI 2233  
 : | | : ||| | : : | | || : | : |||| | : | : |  
 Db 600 VGLGHAARRRLKTYSGMRQRLGLAQALLGEPRLLLLDEPTVGLDPLATVELYQLLDRLR 659  
 Qy 2234 KTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNR 2276  
 | :|| || : | | || : ||| : || : | : :  
 Db 660 GQGTGIVLCSHVLPGVETHIDRAAILAGGRLQVAGSLAELRRK 702

## RESULT 14

US-09-134-000C-6449

; Sequence 6449, Application US/09134000C  
 ; Patent No. 6617156  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al



```
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6449
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6449
```

```
Query Match          2.7%; Score 339.5; DB 4; Length 315;
Best Local Similarity 28.8%; Pred. No. 2.8e-21;
Matches 99; Conservative 61; Mismatches 137; Indels 47; Gaps 7;
```

```
Qy      993 VDKLTKVYKDDKKLALNKLNLNLYENQVVSFLGHNGAGKTTTMSILTGLFPPTSGSATIY 1052
      : | ||| : || : | : | : || ||||:|: |:| | ||| |:
Db      11 IQDLRKVYASGVE-ALRGIDLTVEEGDFYALLGPNGAGKSTTIGIVTSLVNKTSKGVKIF 69

Qy     1053 GHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSRLKSMAQEEIRREMDKMIEDLE 1112
      |:|: ||| :::|: || ||:: : : ::::| : :| ::
Db      70 GYDLDTMVRAKQQIGLVPQEFNPNPFETVQQIVVNQAGYYGVSKEAMKRSEKYLKQSN 129

Qy     1113 LSNKRHSLVQTLSSGGMKRKLSVAIAFVGGSRAIILDEPTAGVDPYARRAIWDLILKYK-P 1171
      | ||: : |||||:| :| | : : :||| |||| | ||:| : :
Db     130 LWEKRNERRARMLSGGMKRRLMIARALMHEPKLLILDEPTAGVDIELRREMWAFLQELNAQ 189

Qy     1172 GRTILLSTHHMDEADLLGDRIAIISHGKLCCKGSPLFLKGTYGDGYRLTLVKRPAEPGGP 1231
      | ||:|:|:|:|:|:| | || |:| | :|
Db     190 GTTIILTTHYLEEAEMLCRNIGIIQSGEL-----IENTSMKH----- 226

Qy     1232 QEPGLASSPPGRAPLSSCSELQVSQFI---RKHVASCLLVSDTST-ELSYILPSEAAKKG 1287
      ::|| || : : :: | | | | :
Db     227 -----LLAKLQFETFIFDLAPYTQAPVIEGYQSVFEDELTLAVEVERNQ 270

Qy     1288 AFERLFQHLERSLDALHLSSFGLMDTTLEEVFLKVSEEDQSLEN 1331
      ||: | | : : | ||:|:|:|:| :|
Db     271 GVNHLFEQL--SQQGIKVLSMRNKSNRLEELFLKITEDTYQRED 312
```

# RESULT 15

```
US-09-252-991A-18351
; Sequence 18351, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
```



Qy 1650 FSCPSSVGGHPPQMRVVTGDILTDITGHNVSEYLLFTSDRFRHLHRYGAITFGNVLSIPA 1709  
 Db 310 -----LADRAQVQ-----LAPVPA 323  
 Qy 1710 SFGTRAPPMVRKIAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGIT 1769  
 Db 324 RF-----EDAFIDLLG-----GGP----- 337  
 Qy 1770 VTNHPMNKTSASLSLDYLLQGTDVVIAIFIIVAMSFVPASFVFLVAEKSTKAKHLQFVS 1829  
 Db 338 ----- 337  
 Qy 1830 GCNPIIYWLANYVWMLNLYLPATCCVILFVFDLPAYTSPTNFPVLSLFLLYGWSITP 1889  
 Db 338 ----- 337  
 Qy 1890 IMYPASFWFEVPSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVVNSYLKSCFLIF 1949  
 Db 338 -----GGTSTLAERL----- 347  
 Qy 1950 PNYNLGHGLMEMAYNEYINEYYAKIGQFDMKSPFEWDIVTRGLVAMAVEGVVGFLLTIM 2009  
 Db 348 ----- 347  
 Qy 2010 CQYNFLRRPQRMVPSTKPVEDDQVASERQVLRGDADNDMVKIENLTKVYKSRKIGRIL 2069  
 Db 348 -----SPVELGSDVA-----VSCRNLTK-----RFGEFT 371  
 Qy 2070 AVDRCLGVRPGECFGLLGVNGAGKTSTFKMLTGDESTTGGEAFVNGHSLVKELLQVQOS 2129  
 Db 372 ATDQVSFEVQKEIFGLLGPNAGKSTTFKMLCGLLKPTAGEAHVVGHDLRHATGAAKSQ 431  
 Qy 2130 LGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYSG 2189  
 Db 432 LGYMAQKFSLYGLLSVRQNLEFSAGVYGLEGNVRRERIEEMIATFDLGDWLSATPDSLPL 491  
 Qy 2190 GNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIKTGRSVVLTSHSMEEC 2249  
 Db 492 GHKQRLALACSLMHRPPVFLDEPTSGVDPITRREFWTHINGLARKGVTIMVTTHFMDEA 551  
 Qy 2250 EALCTRLAIMVNGRLRCLGSIQHLK 2274  
 Db 552 E-YCDRVAMLSRARLIALDTPDALK 575

Search completed: September 1, 2004, 10:58:55  
 Job time : 59 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2004, 10:46:22 ; Search time 68 Seconds  
(without alignments)  
3445.920 Million cell updates/sec

Title: US-10-088-467-2  
Perfect score: 12668  
Sequence: 1 MGFLHQLQLLLWKNVTLKRR.....GLISFEEERAQLSFNTDTLC 2436

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	7979	63.0	1529	2	A59189	ATP-binding casset
2	7119	56.2	1472	2	B54774	ATP binding casset
3	4103	32.4	2201	2	A54774	ATP binding casset
4	2622	20.7	1704	2	S71363	probable ATP-bindi
5	2622	20.7	1704	2	A59188	ATP-binding casset
6	2024	16.0	1802	2	T33783	hypothetical prote
7	1964.5	15.5	1816	2	A84845	probable ABC trans
8	1920	15.2	373	2	T47150	hypothetical prote
9	1718.5	13.6	1447	2	T15200	hypothetical prote
10	1688	13.3	1317	2	C88925	protein F33E11.4 [
11	1524	12.0	1767	2	S60124	transport protein
12	1522.5	12.0	1758	2	F88559	protein C48B4.4b [
13	1515	12.0	1704	2	T42749	ATP-binding casset

14	1448.5	11.4	1246	2	T00826	hypothetical prote
15	1202.5	9.5	1564	2	T27121	hypothetical prote
16	1046	8.3	1431	2	T22748	hypothetical prote
17	746	5.9	269	2	T46467	hypothetical prote
18	707.5	5.6	895	2	T07714	probable ABC-type
19	706.5	5.6	900	2	T07717	probable ABC-type
20	702	5.5	1011	2	T07712	probable ABC-type
21	686	5.4	925	2	T07713	probable ABC-type
22	666	5.3	1336	2	T18288	ABC transport prot
23	649.5	5.1	722	2	T07716	probable ABC-type
24	485	3.8	130	2	I38906	ATP-binding casset
25	469.5	3.7	664	2	T07715	probable ABC-type
26	467.5	3.7	149	2	I38905	ATP-binding casset
27	442.5	3.5	196	2	T12512	hypothetical prote
28	427.5	3.4	327	2	D72257	hypothetical prote
29	421.5	3.3	260	2	T15237	hypothetical prote
30	417	3.3	324	2	C71081	probable resistanc
31	415.5	3.3	350	2	B69065	ABC transporter (A
32	411.5	3.2	328	2	E75108	daunorubicin resis
33	408	3.2	339	2	S74048	probable daunorubi
34	405	3.2	330	2	S27707	daunorubicin resis
35	404.5	3.2	347	2	S76278	ABC-type transport
36	402.5	3.2	300	2	AG2116	ABC transporter AT
37	402	3.2	311	2	G69803	ABC transporter (A
38	402	3.2	333	2	D72492	probable ABC trans
39	398	3.1	275	2	D90267	ABC transporter, A
40	395	3.1	310	2	E96920	ABC transporter (A
41	394	3.1	325	2	S32908	hypothetical prote
42	394	3.1	727	2	T07718	probable ABC-type
43	393.5	3.1	312	2	C69012	ABC transporter (A
44	392.5	3.1	297	2	AE1816	ABC transporter (A
45	391	3.1	305	2	E75122	hypothetical prote

#### ALIGNMENTS

#### RESULT 1

A59189

ATP-binding cassette transporter - human (fragment)

N;Alternate names: KIAA1062 protein

C;Species: Homo sapiens (man)

C;Date: 18-Feb-2000 #sequence\_revision 18-Feb-2000 #text\_change 02-Jun-2000

C;Accession: A59189

R;Kikuno, R.; Nagase, T.; Ishikawa, K.; Hirose, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.

DNA Res. 6, 197-205, 1999

A;Title: Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.

A;Reference number: Z22961; MUID:99397452; PMID:10470851

A;Accession: A59189

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-1529 <KIK>

A;Cross-references: GB:AB028985; NID:g5689460; PIDN:BAA83014.1; PID:d1046841; PID:g5689461

A;Experimental source: chromosome 9; clone hj03579; clone lib pBluescriptII SK plus; tissue type brain.

C;Genetics:

A;Map position: 9

A;Note: KIAA1062

C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 63.0%; Score 7979; DB 2; Length 1529;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1529; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	908	GILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSVMEEDQAC	967
Db	1	GILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSVMEEDQAC	60
Qy	968	AMESRRFEETRGMEEEPHTLPLVVCVDKLTKVYKDDKKLALNKLSLNLYENQVVSFLGHN	1027
Db	61	AMESRRFEETRGMEEEPHTLPLVVCVDKLTKVYKDDKKLALNKLSLNLYENQVVSFLGHN	120
Qy	1028	GAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLW	1087
Db	121	GAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLW	180
Qy	1088	FYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAFVGGSRAIIL	1147
Db	181	FYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAFVGGSRAIIL	240
Qy	1148	DEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPL	1207
Db	241	DEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPL	300
Qy	1208	FLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHAVASCLL	1267
Db	301	FLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHAVASCLL	360
Qy	1268	VSdTSTELSYILPSEAAKKGAferLFQHLERSLDALHLSSFGLMDTTLEEVLKVSEEDQ	1327
Db	361	VSdTSTELSYILPSEAAKKGAferLFQHLERSLDALHLSSFGLMDTTLEEVLKVSEEDQ	420
Qy	1328	SLENSEADVKEsrKdVLPgaEGPASGEGHAGNLARCSELTQSQASLQSASSVGSARGDEG	1387
Db	421	SLENSEADVKEsrKdVLPgaEGPASGEGHAGNLARCSELTQSQASLQSASSVGSARGDEG	480
Qy	1388	AGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVQGGSRKLDGGWLKVRQFHGLLVKR	1447
Db	481	AGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVQGGSRKLDGGWLKVRQFHGLLVKR	540
Qy	1448	FHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYHNYTQPRGNFIP	1507
Db	541	FHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYHNYTQPRGNFIP	600
Qy	1508	YANEERREYRLRLSPDASPQQILVSTFRLPSGVGATCVLKSPANGSLGPTLNLSGSESRL	1567
Db	601	YANEERREYRLRLSPDASPQQILVSTFRLPSGVGATCVLKSPANGSLGPTLNLSGSESRL	660
Qy	1568	AARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDELQAWNVS LPPTAGPEMWT	1627

Db	661		AARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDQLQAWNVS LPPTAGPEMWT	720
Qy	1628		APSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILT DITGHNVS EYLLFTS	1687
Db	721		APSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILT DITGHNVS EYLLFTS	780
Qy	1688		DRFRLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSMPTYLNSL	1747
Db	781		DRFRLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSMPTYLNSL	840
Qy	1748		NNAILRANLPKSKGNPAAYGITVTNHMPNKTSASLSLDYLLQGT DVVIAIFIIVAMS FVP	1807
Db	841		NNAILRANLPKSKGNPAAYGITVTNHMPNKTSASLSLDYLLQGT DVVIAIFIIVAMS FVP	900
Qy	1808		ASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVII LFVFDLPAY	1867
Db	901		ASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVII LFVFDLPAY	960
Qy	1868		TSPTNFPVAVLSLFLLYGWSITPIMYPASFWFVPS SAYVFLIVINLFIGITATVATFLLQ	1927
Db	961		TSPTNFPVAVLSLFLLYGWSITPIMYPASFWFVPS SAYVFLIVINLFIGITATVATFLLQ	1020
Qy	1928		LFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFEWD	1987
Db	1021		LFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFEWD	1080
Qy	1988		IVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDD VDVASERQ RVLRGDAD	2047
Db	1081		IVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDD VDVASERQ RVLRGDAD	1140
Qy	2048		NDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLG VNGAGKTSTFKMLTGDEST	2107
Db	1141		NDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLG VNGAGKTSTFKMLTGDEST	1200
Qy	2108		TGGEAFVNGH SVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVV	2167
Db	1201		TGGEAFVNGH SVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVV	1260
Qy	2168		KWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWN	2227
Db	1261		KWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWN	1320
Qy	2228		LILDLIKTRGSVVLTSHSMECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVRT	2287
Db	1321		LILDLIKTRGSVVLTSHSMECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVRT	1380
Qy	2288		KSSQSVKDVVRFFNRNFP EAM LKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDY	2347
Db	1381		KSSQSVKDVVRFFNRNFP EAM LKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDY	1440
Qy	2348		SVSQTTLDNVFNFAKKQSDNLEQQETEP PSALQSPLGCLLSLLRPRSAPTELRLVADE	2407
Db	1441		SVSQTTLDNVFNFAKKQSDNLEQQETEP PSALQSPLGCLLSLLRPRSAPTELRLVADE	1500
Qy	2408		PEDLDTEDEGLISFEEERAQLSFNTDTLC	2436

Db 1501 PEDLDTEDEGLISFEEERAQLSFNTDTLC 1529

RESULT 2

B54774

ATP binding cassette transporter ABC2 - mouse (fragment)

C;Species: Mus musculus (house mouse)

C;Date: 23-Mar-1995 #sequence\_revision 05-Apr-1995 #text\_change 02-Feb-2001

C;Accession: B54774

R;Luciani, M.F.; Denizot, F.; Savary, S.; Mattei, M.G.; Chimini, G.

Genomics 21, 150-159, 1994

A;Title: Cloning of two novel ABC transporters mapping on human chromosome 9.

A;Reference number: A54774; MUID:94375008; PMID:8088782

A;Accession: B54774

A;Molecule type: mRNA

A;Residues: 1-1472 <LUC>

A;Cross-references: GB:X75927; NID:g495258; PIDN:CAA53531.1; PID:g495259

C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

C;Keywords: ATP; nucleotide binding; P-loop

F;44-234/Domain: ATP-binding cassette homology <ABC1>

F;61-68/Region: nucleotide-binding motif A (P-loop)

F;1108-1300/Domain: ATP-binding cassette homology <ABC2>

F;1126-1133/Region: nucleotide-binding motif A (P-loop)

Query Match 56.2%; Score 7119; DB 2; Length 1472;

Best Local Similarity 94.2%; Pred. No. 0;

Matches 1388; Conservative 22; Mismatches 60; Indels 4; Gaps 4;

```
Qy      965 QACAMESRRFEETRGMEEPTHLPVVCVDKLTKVYKDDKKLALNKLNLNLYENQVVSFL 1024
          |||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db      1   QACAMESRHFEETRGMEEPTHLPVVCVDKLTKVYKNDKKLALNKLNLNLYENQVVSFL 60

Qy     1025 GHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEE 1084
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      61 GHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEE 120

Qy     1085 HLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSIVQTLSSGGMKRKLSVAIAFVGGSGRA 1144
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db     121 HLWFYSRLKSMAQEEIRKETDKMIEDLELSNKRHSIVQTLSSGGMKRKLSVAIAFVGGSGRA 180

Qy     1145 IILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKCLKCCG 1204
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db     181 IILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKCLKCCG 240

Qy     1205 SPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHHVAS 1264
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db     241 SPLFLKGAYXDGYRLTLVKQPAEPGTSQEPGLASSPSGCPRLSSCSEPQVSQFIRKHHVAS 300

Qy     1265 CLLVSDTSTELSYILPSEAAKKGAFERLFQHLERSLDALHLSSFGLMDTTLEEVFLKVSE 1324
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db     301 SLLVSDTSTELSYILPSEAVKKGAFERLFQQLEHSLDALHLSSFGLMDTTLEEVFLKVSE 360

Qy     1325 EDQSLENSEADVKESSRKDVLPGAEGPASGEGHAGNLARCSELTQSQASLQSASSVGSARG 1384
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db     361 EDQSLENSEADVKESSRKDVLPGAEGLTAVGGQAGNLARCSELAQSQASLQSASSVGSARG 420
```



Qy	1385	DEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVQGQSRKLDGGWLKVRQFHGLL	1444
		:     :                        :	
Db	421	EEGTGYSDGYGDYRPLFDNLQDPDNVSLQEAEAMEALAQVGQSRKLEGWWLKMRQFHGLL	480
Qy	1445	VKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYHNYTQPRGN	1504
Db	481	VKRFHCARRNSKALCSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYHNYTQPRGN	540
Qy	1505	FIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGES	1564
Db	541	FIPYANEERQEYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGES	600
Qy	1565	RLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPED-LQAWNVLSPPTAGPE	1623
Db	601	RLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPVXPDEDSLQAWNMSLPPTAGPE	660
Qy	1624	MWTSAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTDTGHNVSEYL	1683
Db	661	TWTSAPSLPRLVHEPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTDTGHNVSEYL	720
Qy	1684	LFTSDRFLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSMPTY	1743
Db	721	LFTSDRFLHRYGAITFGNVQKSIPASFGARVPPMVRKIAVRRVAQVLYNNKGYHSMPTY	780
Qy	1744	LNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFIIVAM	1803
Db	781	LNSLNNAILRANLPKSKGNPAAYXITVTNHPMNKTSASLSLDYLLQGTDVVIAIFIIVAM	840
Qy	1804	SEVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWMLNYLVPATCCVILFVFD	1863
Db	841	SEVPASFVFLVAEKSTKAKHLQFVSGCNPVIYWLANYVWMLNYLVPATCCVILFVFD	900
Qy	1864	LPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITATVAT	1923
Db	901	LPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITATVAT	960
Qy	1924	FLLQLFEHDKDLKVVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSP	1983
Db	961	FLLQLFEHDKDLKVVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSP	1020
Qy	1984	FEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDQDVASERQRVLR	2043
Db	1021	FEWDIVTRGLVAMTVEGFVGFLLTIMCQYNFLRQPQRLPVSTKPVEDDQDVASERQRVLR	1080
Qy	2044	GDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGV-RPGECFGLLVNGAGKTSTFKMLT	2102
Db	1081	GDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVCVPGECEGGLLVNGAGKTSTFKMLT	1140
Qy	2103	GDESTTGGEAFVNGHSLVKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKD	2162
Db	1141	GDESTTGGEAFVNGHSLVKDLLQVQQSLGYCPQFDPVDELTAAREHLQLYTRLRCIPWKD	1200
Qy	2163	EARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKAR	2222
		:	
Db	1201	EAQVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKAR	1260
Qy	2223	RFLWNLIILDIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYM	2282

```

Db      1261 RFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLHCLGSIQHLKNRFGDGYM 1320
Qy      2283 ITVRTKSSQSVKDVVRFFNRFNPFPEAMLERHHTKVQYQLKSEHISLAQVFSKMEQVSGVL 2342
        |||||:|||||:|:| ||||| |||
Db      1321 ITVRTKSSQNVKDVVRFFNRFNPFPEAHQGKTPYKVQYQLKSEHISLAQVFSKMEQVVGVL 1380
Qy      2343 GIEDYSVSQTTLDNVFNFAKKQSDNLEQQETEPPSALQSPLGCLLSLLRPRSAPTELRA 2402
        |||||:||||| ||:| ||| ||||| |||||
Db      1381 GIEDYSVSQTTLDNVFNFAKKQSDNVEQQEAE-PSSLPSPLG-LLSLLRPRPAPTELRA 1438
Qy      2403 LVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2436
        |||||:||||| |||||
Db      1439 LVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 1472

```

# RESULT 3

A54774

ATP binding cassette transporter ABC1 - mouse

C;Species: Mus musculus (house mouse)

C;Date: 05-Apr-1995 #sequence\_revision 05-Apr-1995 #text\_change 02-Feb-2001

C;Accession: A54774

R;Luciani, M.F.; Denizot, F.; Savary, S.; Mattei, M.G.; Chimini, G.

Genomics 21, 150-159, 1994

A;Title: Cloning of two novel ABC transporters mapping on human chromosome 9.

A;Reference number: A54774; MUID:94375008; PMID:8088782

A;Accession: A54774

A;Molecule type: mRNA

A;Residues: 1-2201 <LUC>

A;Cross-references: GB:X75926; NID:g495256; PIDN:CAA53530.1; PID:g495257

C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

C;Keywords: ATP; duplication; nucleotide binding; P-loop

F;856-1047/Domain: ATP-binding cassette homology <ABC1>

F;873-880/Region: nucleotide-binding motif A (P-loop)

F;1869-2060/Domain: ATP-binding cassette homology <ABC2>

F;1886-1893/Region: nucleotide-binding motif A (P-loop)

Query Match 32.4%; Score 4103; DB 2; Length 2201;

Best Local Similarity 40.9%; Pred. No. 2.1e-263;

Matches 943; Conservative 323; Mismatches 626; Indels 414; Gaps 58;

```

Qy      244 SGELGRILTVPES-----QKGALQGYR---DAVCSGQAAARARRFSGLSAELR 288
        || | |::| | || |||: :::| |
Db      95 SGFLQHNLSLPRSTVDSLLQXNVGLQKVFLQGYQLHLASLCNGS-----KLEEI 143
Qy      289 NQLDVAKVSQQGLGLDAPNGSDSSPQAPPPRRLLQALLGDLLDAQKVLQ-DVDVLSALALL 347
        || |::| || |::| | |::||: ::||: |
Db      144 IQLGDAEVSALCGL-----PRKKLDA-----AERVLRYNMDILKPVVTKL 183
Qy      348 PQGACTGRTPGPPASGA-----GGAAN---GTGAGAVMGPNATAEEGAPSAAALATP 396
        | | :| || | |::| |
Db      184 NS---TSHLPTQHLAEATTVLLDSLGLAQELFSTKSWSDMRQEVMTNVTNNS----- 234
Qy      397 DTLQGQCSAFVQLWAGLQPILCGNNRTIEPE-----ALRRGN----- 433
        |: |:: : |::| || ||
Db      235 -----SSSTQIYQAVSRIVCGH-----PEGGGLKIKSLNWDYEDNNYKALFGGNNTEED 282

```

Qy	434	-----MSSLGFTSKEQRNLGLLVHLMTSNP-----KILYAPAGSEVDRV	472
Db	283	VDTFYDNSTTPYCNDLMKNL-----ESSPLSRIIWKALKPLLVGKILYTPDTPATRQV	335
Qy	473	ILKANETFAFVGNVTHYAQVWLNISAEIRSFLEQGRLLQHLRWL-----	516
Db	336	MAEVNKTFOELAVFHDLEGMWEELSPOIWTFMENSQEMDLVRTLLDSRGNDQFWEQKLDG	395
Qy	517	-----QQYVAELRLHPEAL---NLSLDELPPALRQDNFSLPSGMALLQQLDITDNAACGW	568
Db	396	LDWTAQDIMAFIAKPNEDVQSPNGSVYTWREAFNETN-----QAIQTIS-----	439
Qy	569	IQFMSKVSVDIFKGFPEDEESIVNYTLNQAYQDNVTVFASVIFQ--TRKDGSLPPHVHYKI	626
Db	440	-RFMECVNLNKLPIPIEVRLINKSME--LLDERKFWAGIVFTGITPDSVELPHHVYKI	496
Qy	627	RQNSSFTEKTNEIRRAYWRPGPNTG---GRFYFLYGFVWIQDMMERAIIDTFVGHADVPEP	683
Db	497	RMDIDNVERTNKIKDGYWDPGRADPFEDMRVWGGFAYLQDVVEQAIIRVLTGSE-KKT	555
Qy	684	GSYVQMFPPYPCYTRDDFLFVIEHMMPLCMVISWVYSVAMTIQHIVAEKEHRLKEVMKTMG	743
Db	556	GVYVQQMPYPCYVDDIFLRVMSRSMPLFMTLAWIYSVAVIKSIYVEKEARLKETMRIMG	615
Qy	744	LNNAVHWWAVFITGFVQLSISVTALTALIKYGQVLMHSHVVIWFLAVYAVATIMFCFL	803
Db	616	LDNGILWFSWFVSSLIPLLVSAGLLVILKLGNLPLPYSDPSVVFVFLSVFAMVTILQCFL	675
Qy	804	VSVLYSKAKLASACGGIIFYLSYVPMYVAIREEVAHDKITAFE-KCIASLMSTTAFGLG	862
Db	676	ISTLFSRANLAAACGGIIFYTLPLPYVLC-----VAWQDYVGFSIKIFASLLSPVAFGFG	730
Qy	863	SKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTMLMVDVAVYGILTWYIEAVHPGMY	922
Db	731	CEYFALFEEQIGVQWDNLFESPVEEDGFNLTTAVSMMLFDTFLYGVMTWYIEAVFPQY	790
Qy	923	GLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSVMEEDQACAMESRRFEETRGMEE	982
Db	791	GIPRPWYFPCTKSYWFG---EIDEKSHPGSSQKGV-----EIC-----MEE	830
Qy	983	EPHLPLVVCVDKLTKVYKDDKKLALNKLSLNLYENQVVSFLGHNGAGKTTTMSILTGLF	1042
Db	831	EPHLRLGVSIQNLVKVYRDGMKVAVDGLALNFYEGQITSFLGHNGAGKTTTMSILTGLF	890
Qy	1043	PPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSRLKSMAQEEIRR	1102
Db	891	PPTSGTAYILGKDIRSEMSSIRQNLGVCPQHNVLFDMLTVEEHIWFYARLKGLSEKHVKA	950
Qy	1103	EMDKMIEDLEL-SNKRHSLVQTLSGGMKRKLSVAIAFVGGSRAIILDEPTAGVDPYARRA	1161
Db	951	EMEQMALDVGLPPSKLSKTSQLSGGMQRKLSVALAFVGGSKVVLIDEPYARRA	1010
Qy	1162	IWDLILKYKPGRTIILLSTHMHDEADLLGDRIAIISHGKLKCCGSPFLKGTGYDGYRLTL	1221
Db	1011	IWELLLKYRQGRTIILLSTHMHDEADILGDRIAIISHGKLCCVGSSFLKNQLGTGYLTL	1070

Qy 1222 VKRPAEPG-----GPQEPGLASSPPGRAPLSSCSELQVSQFIR 1259  
 ||: | : || | | :| ||  
 Db 1071 VKKDVESSLSSCRNSSSTVSLCKKEDSVSQSSSDAGLGSDHESDTLTIDVS--AISNLIR 1128

Qy 1260 KHVASCLLVSDTSTELSYILPSEAAKKGAFLERLFQHLERSLDALHLSSFGLMDTTLEEVF 1319  
 |||: || | ||:|:| |||:| || :| | :||:| :|||:|  
 Db 1129 KHVSEARLVEDIGHELTYVLPYEAAKEGAFVELFHEIDRLSDLGISSYGISETTLEEIF 1188

Qy 1320 LKVSEEDQSLENSEADVKEARKDVLPGAEGPASGEGHAGNLARCSELTQSQASLQSASSV 1379  
 |||:| | | | : ||  
 Db 1189 LKVAEE-----SGVDA-ETSDGTLF----- 1207

Qy 1380 GSARGDEGAGYTDVYGDYRPLF-----DNPQDPD--NVSLEVEAEALSRV-GQGSRKLD 1431  
 || : | :|| : | : ||: : : | : || : | :| :|  
 Db 1208 --ARRNRR-----FGDKQSCSLHPFTEDDAVDPNDSIDPESRETDLGSGMDGKGSYQLK 1260

Qy 1432 GGWLKVRQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVP EIGDLPPLVLS 1491  
 | | :|| || || |||: | | :||:| |||:| :| || | | | |  
 Db 1261 GWKLTQQQFVALLWKRLLIARRSRKGFFAQIVLPAVFCIALVFSLIVPPFGKYPSELEQ 1320

Qy 1492 PSQYH-NYTQPRGNFI PYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPAN 1550  
 | | : || : : : | | | :| : | | | : :|  
 Db 1321 PWMYNEQYT-----FVSNDAP-----DMGTQELLNALT KDPGFGTRCMEGNPIP 1365

Qy 1551 GSLGPTLNLSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPEDLQ 1610  
 | : || : :||  
 Db 1366 -----DTPCL-----AGEED-- 1375

Qy 1611 AWNVSLPPTAGPEM-----WTSAPSLPRLVREPVRCTCSAQGTGFS---CPSSVGG-HPP 1661  
 | :| | : :| || | |||: || || ||  
 Db 1376 -WTISPVPQSIVDLFQNGNWTMKNPSP-----ACQCSSDKIKKMLPVCPPGAGGLPPP 1427

Qy 1662 QMRVVTGDILTDITGHNVSLEYLLFTSDRF-----RLHRYGAITFG----- 1701  
 | : | ||| :|| | :|:| : | : ||| :|  
 Db 1428 QRKQKTADILQNL TGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNSQALPP 1487

Qy 1702 -----NVLKSIPASFGTRAPPMVRKIA-----VRRAAQVFYNNKGYHSMPTYLN 1745  
 : | : : | | : : : :|||:| :| :|  
 Db 1488 SHEVNDAIKQMKLLKLT KDTSADRFLSSLGRFMAGLDTKNNVKVWFNNKGWHAISSFLN 1547

Qy 1746 SLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLS-LDYLLQGTDVVIAIFIIVAMS 1804  
 :||| ||| | : ||: ||| |||:| | || : : |||:| :| |||  
 Db 1548 VINNAILRANLQKGE-NPSQYGITAFNHPLNLTKQQLSEVALMTTSVDVLVSICVIFAMS 1606

Qy 1805 FVPASFVVFVLAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVIIIFVFDL 1864  
 ||| ||| : | :||| |||:| | :|||:| :||| ||| :|| |  
 Db 1607 FVPASFVFLIQERVSKAKHLQFISGVKPVYIYWLSNFVWDMCNYVVPATLVIIIFICFQQ 1666

Qy 1865 PAYTSPTNFPAVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGITATVATF 1924  
 :| | || | : | ||| |||:| ||| | :|||:| | :||| :|||  
 Db 1667 KSYVSSSTNLPVLALLLLLYGWSITPLMYPASFVKIPSTAYVVLTSVNLFIGINGSVATF 1726

Qy 1925 LLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPF 1984  
 :||| :| | :| ||| |||:| || |||:| | : : : | : : ||  
 Db 1727 VLELFTNNK-LNDINDILKSVFLIFPHFCLGRGLIDMVKNQAMADALERFGE-NRFVSP 1784

Qy 1985 EWDIVTRGLVAMAVEGVVGLLTIMCQYNFLRRPQRMVPSTKPVED-DVDVASERQRVLR 2043

```

      ||:| | | ||||| ||:|: || | ||: : | : | | | ||||:|
Db      1785 SWDLVGRNLFAMAVEGVVFFLITVLIQYRFFIRPRPVKAKLPPLNDEDEDVRRERQRILD 1844

QY      2044 GDADNDMVKIENLTKVYKSRKIGRILAVDRCLGVPRGECFGLLGVNGAGKTSTFKMLTG 2103
      | ||:|:|: |||:|: : | ||||:|:|: ||||| |||||:|:| |||||
Db      1845 GGGQNDILEIKELTKIYRRK---RKPAVDRICIGIPPGECEFGLLGVNGAGKSTTFKMLTG 1901

QY      2104 DESTTGGEAFVNGHVSVLKELLQVQQSLGYCPCDALFDELTAREHLQLYTRLRGISWKDE 2163
      | | |:|:| :|:| : :| |:| |||| ||: : || |||: : |||: |:
Db      1902 DTPVTRGDAFLNKNLSILSNIHEVHQNMGYCPQFDAITELLTGREHVEFFALLRGVPEKEV 1961

QY      2164 ARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARR 2223
      : |:|: || | || :| | ||||| |||||:| ||| | :||| |||||
Db      1962 GKFGEWAIRKLGLVKYGEKYASNYSGGNKRKLSTAMALIGGPPVFLDEPTTGMDPKARR 2021

QY      2224 FLWNLIILDIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMI 2283
      |||| | :| | ||||| |||||:| ||||| |||||:| ||||| |
Db      2022 FLWNCALSIVKEGRSVVLTSHSMEECEALCTRM AIMVNGRFRCLG SVQHLKNRFGDGYTI 2081

QY      2284 TVR-TKSSQSVKDVVRFFNRFPEAMLEKRRHHTKVQYQLKSEHISLAQVFSKMEQVSGVL 2342
      || |: :| | || || :|:|:| :||| | |||:| | :| |
Db      2082 VVRIAGSNPDLKPVQEFFGLAFPGSVLKEKHRNMLQYQLPSSLSSLARIFSILSQSKKRL 2141

QY      2343 GIEDYSVSQTTLDNVFVNFAKKQSDN 2368
      ||||| ||||| |||:
Db      2142 HIEDYSVSQTTLDQVFNFAKDQSD 2167

```

#### RESULT 4

S71363

probable ATP-binding cassette transporter ABC-3 - human

N;Alternate names: ATP-binding cassette transporter ABC-C

C;Species: Homo sapiens (man)

C;Date: 29-Jan-1998 #sequence\_revision 06-Feb-1998 #text\_change 02-Feb-2001

C;Accession: S71363

R;Klugbauer, N.; Hofmann, F.

FEBS Lett. 391, 61-65, 1996

A;Title: Primary structure of a novel ABC transporter with a chromosomal localization on the band encoding the multidrug resistance-associated protein.

A;Reference number: S71363; MUID:96326608; PMID:8706931

A;Accession: S71363

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-1704 <KLU>

A;Cross-references: EMBL:X97187; NID:g1514529; PIDN:CAA65825.1; PID:e243436; PID:g1514530

A;Experimental source: cell line medullary thyroid carcinoma

C;Genetics:

A;Gene: GDB:ABC3

A;Cross-references: GDB:3770735; OMIM:601615

A;Map position: 16p13.3-16p13.3

C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

C;Keywords: ATP binding; nucleotide binding; P-loop; phosphoprotein; transmembrane protein

F;255-283/Domain: transmembrane #status predicted <TM1>

F;307-329/Domain: transmembrane #status predicted <TM2>

F;345-364/Domain: transmembrane #status predicted <TM3>  
 F;373-394/Domain: transmembrane #status predicted <TM4>  
 F;401-422/Domain: transmembrane #status predicted <TM5>  
 F;452-475/Domain: transmembrane #status predicted <TM6>  
 F;549-739/Domain: ATP-binding cassette homology <ABC1>  
 F;566-573/Region: nucleotide-binding motif A (P-loop)  
 F;685-690/Region: nucleotide-binding motif B  
 F;1100-1120/Domain: transmembrane #status predicted <TM7>  
 F;1145-1169/Domain: transmembrane #status predicted <TM8>  
 F;1181-1207/Domain: transmembrane #status predicted <TM9>  
 F;1215-1236/Domain: transmembrane #status predicted <TM10>  
 F;1245-1264/Domain: transmembrane #status predicted <TM11>  
 F;1299-1324/Domain: transmembrane #status predicted <TM12>  
 F;1399-1590/Domain: ATP-binding cassette homology <ABC2>  
 F;1416-1423/Region: nucleotide-binding motif A (P-loop)  
 F;1535-1540/Region: nucleotide-binding motif B  
 F;674,866,1524/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status predicted  
 F;1344/Binding site: phosphate (Thr) (covalent) (by cAMP-dependent kinase) #status predicted

Query Match 20.7%; Score 2622; DB 2; Length 1704;  
 Best Local Similarity 34.0%; Pred. No. 4.2e-165;  
 Matches 638; Conservative 317; Mismatches 556; Indels 364; Gaps 45;

Qy	581	KGFPDEESIVNYTLNQAYQDNV--TVFASVIFQ---TRKDGSLLPPHVHYKIR-----	627
		:     : :     :   : : :        :	
Db	108	RGFPSEKDFEDY----IRYDNCSSSVLAADVFEHPFNHSKEPLPLAVKYHLRFSYTRRNY	163
Qy	628	---QNSSFTEK-----TNEIRRAYWRPG-----PNTGGRFYFLYGFWIQDMMERAI	672
		: :     :       : :  ::  :	
Db	164	MWTQTGSFFLKETEGWHTTSLFPLFPNPGPREPTSPDGGEPGYIREGFLAVQHAVDRAIM	223
Qy	673	DTFVGHDVVEPGSY-----VQMFYPYCYTRDDFLFVIEHMMPLCMVISWVYSVAMTIQH	726
		:   : ::     :       : :   :: :   :	
Db	224	EYHA--DAATRQLFQRLTVTIKRFYPPIADPFLVAIQYQLPLLLLLSFTYTALTIAIRA	281
Qy	727	IVAEKEHRLKEVMKTMGLNNAVHWVAFITGFVQLSISVTALTAIL-----KYGQVLMHS	781
		:           :    : :      :  :    : : :  :	
Db	282	VVQEKERRLKEYMRMMGLSSWLHWSAWFLFLFLIAASEMTLLFCVKVKPNVAVLSRS	341
Qy	782	HVVIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHD	841
		::    : :        :    : :     :   : :  :     ::	
Db	342	DPSLVLAFLLCFAISTISFSFMVSTFFSKANMAAAFGGFLYFFTYIPYFFVAPR----YN	397
Qy	842	KITAFEKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVE-GDDFNLLLAVTML	900
		:  :  :  :    : : :   :            :	
Db	398	WMTLSQKLCSCLLSNVAMAMGAQLIGKFEAKMGIQWRDL-LSPVNVDDDFCFGQVLGML	456
Qy	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSV	960
		:: : : : : : :      : : : :  :           :	
Db	457	LLDSVLYGLVTWYMEAVFPGQFGVPQWPYFFIMPYSWCGKPRAVAGK-----	503
Qy	961	MEEDQACAMESRRFEETRGMEEPTHLPVVCVDKLTKVYK--DDKKLALNKLNLNLYEN	1018
		: : : :          : :  : : : : :  :  :	
Db	504	-EEEDSDPEKALRNEY---FEAEPEDLVAGIKIKHLSKVFRVGNKDRAAVRDLNLNLYEG	559

Qy	1019	QVVSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLF	1078
Db	560	QITVLLGHNGAGKTTTLSMLTGLFPPTSGRAYISGYEISQDMVQIRKSLGLCPQHDILFD	619
Qy	1079	RLTVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAF	1138
Db	620	NLTVAEHLFYAQLKGLSRQKCPEEVKQMLHIIGLEDKWNRSRFLSGGMRRKLSIGIAL	679
Qy	1139	VGGSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHG	1198
Db	680	IAGSKVLILDEPTSGMDAISRRAIWDLQKQSDRTIVLTTHFMDEADLLGDRIAIMAKG	739
Qy	1199	KLKCCGSPLFLKGTYG DYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFI	1258
Db	740	ELQCCGSSLFLKQKYGAGYHMTLVKEP-----HCNPEDISQLV	777
Qy	1259	RKHVASCLLVSDTSTELSYILPSEAAKKGAFERLFQHLERSLDALHLSSFGLMDTTLLEV	1318
Db	778	HHHVPNATLESSAGAELS FILPRESTR--FEGLEFAKLEKKQKELGIA SFGASITTMEEV	835
Qy	1319	FLKVSEEDQSLENSEADV KESRKDVLPGAEGPASGEGHAGN LARCS ELTQSQASLQSASS	1378
Db	836	FLRVGK---LVDSSMDIQAIQ---LPALQ--YQHERRASDWAVDSNL-----	874
Qy	1379	VGSARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVQGSRKLDGGW-LKV	1437
Db	875	CGAMDPSDGIG-----ALIEEERTAV-----KLNTGLALHC	905
Qy	1438	RQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVP EIGDLPPLVLSPSQYHN	1497
Db	906	QQFWAMFLKKAAYS WREWKMVAAQVLVPLTCVTLALLAINYSSELFDDPMLRLTLGEY--	963
Qy	1498	YTQPRGNFIPYANEERREYRLRLSPDASPOQLVSTFRLP SGVGATCV-LKSPANGSLGPT	1556
Db	964	-----GRTVVPFVSVPGTSQLGQQ	981
Qy	1557	LNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPD EDLQAWNVS L	1616
Db	982	LS-----	983
Qy	1617	PPTAGPEMWT'SAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTDITG	1676
Db	984	-----EHLKDALQAE-----QEPREVLGDL-----	1004
Qy	1677	HNVS EYLLFTSDRFR LHRYGAITFG--NVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNN	1734
Db	1005	---EEFLIFRA-----SVEGGGFNERCLVAASF-----RDVGERTVVNALFNN	1044
Qy	1735	KGYHSMPTYLNSILNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQG-TDV	1793
Db	1045	QAYHSPATALAVVDNLLF-----KLLCGPHA-SIVVS NFQPR SALQA AKDQFNEGRKGF	1098
Qy	1794	VIAIFIIVAMSFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMLNLYLPAT	1853
Db	1099	DIALNLLFAMAFLASTFSILAVSERAVQAKHVQFVSGVHVASFWSALLWDLISFLIPSL	1158
Qy	1854	CCVIILFVFDLPAYTSPTNFP AVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINL	1913

```

      ::: ||: |:| : | | ||||:| |:| |:| ::|| | : |:
Db      1159 LLLVVFKAQDRAFTRDGHMADTLLLLLLYGWAIIPLMYLMNFFFLGAATAYTRLTIFNI 1218

Qy      1914 FIGITATVATFLLQLFEHDKDLKV--VNSYLKSCFLIFPNYNLGHGLMEMAYNEY----- 1966
      || ||||: |:| :: | ||: ||: || : | |
Db      1219 LSGI----ATFLMV'TIMRIPAVKLEELSKTLDHVFLVLPNHCLGMAVSSF-YENYETRRY 1273

Qy      1967 -----INEYYAKIGQFDKMKSPFEWDI--VTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQ 2019
      : :| | :: : | | | : || | :| : : | |:| :
Db      1274 CTSSEVAAHYCKYNIQYQENFYAWSAPGVGRFVASMAASGCAYLILLFLIETNLLQRLR 1333

Qy      2020 -----RMPVSTKPVEDDQDVASERQVLRGDADNDM---VKIENLTKV 2059
      |||| : :| ||| || |:| |: : : |: |:|
Db      1334 GILCALRRRRTLTELYTRMPV----LPEDQDVADERTRILAPSPDLLHTPLIIKELSKV 1389

Qy      2060 YKSRKIGRILAVDRLCLGVRPGECFGLLGVNGAGKTSTFKMLTGDESTTGGEAFVNGHSV 2119
      |: | :||| | |: ||||| |||||:|||||:| | |:| | | :
Db      1390 YEQRV--PLLAVDRSLAVQKGEFCFGLLGFGAGKTTTFKMLTGEESLTSGDAFVGGHRI 1447

Qy      2120 LKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKY 2179
      :: :|:| :||| | || | :| || | :| |||| : | : | | | :
Db      1448 SSDVGKVRQRIGYCPQFDALLDHMTGREMLVMYARLRGIPERHIGACVENTLRGLLLEPH 1507

Qy      2180 ADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIKTGRSV 2239
      |:| ||||| |||| | || | ||||:|||| | || | : : :|:|:|
Db      1508 ANKLVRTYSGGNKRKLSTGIALIGEPAVIFLDEPSTGMDPVARRLLWDTVARARESGKAI 1567

Qy      2240 VLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVRTKS---SQSVKDV 2296
      ::||| ||||| |||| | : :||| ||||:| | | : : :| : : : :
Db      1568 IITSHSMEECEALCTRLAIMVQGQFKCLGSPQHLKSKFGSGYSLRAKVQSEGQQAEEF 1627

Qy      2297 VRFFNRFPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLN 2356
      | : || :|:| | | | | :| |:| |:| |:| ||||| |:|
Db      1628 KAFVDLTFPGSVLEDEHQGMVHYHLPGRDLWAKVFGILEKAKEKYGVDDYSVSQISLEQ 1687

Qy      2357 VEVNFQAKQSDNLEQ 2371
      ||:| | | :
Db      1688 VFLSFAHLQPPTAEE 1702

```

# RESULT 5

A59188

ATP-binding cassette transporter ABC3 - human

C;Species: Homo sapiens (man)

C;Date: 18-Feb-2000 #sequence\_revision 18-Feb-2000 #text\_change 17-May-2002

C;Accession: A59188

R;Connors, T.D.; van Raay, T.J.; Petry, L.R.; Klinger, K.W.; Landes, G.M.; Burn, T.C.

Genomics 39, 231-234, 1997

A;Title: The cloning of a human ABC gene (ABC3) mapping to chromosome 16p13.3.

A;Reference number: A59188; MUID:97179225; PMID:9027511

A;Accession: A59188

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-1704 <CON>

A;Cross-references: GB:U78735; NID:g1699037; PIDN:AAC50967.1; PID:g1699038

C;Genetics:



C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 20.7%; Score 2622; DB 2; Length 1704;  
Best Local Similarity 34.0%; Pred. No. 4.2e-165;  
Matches 638; Conservative 317; Mismatches 556; Indels 364; Gaps 45;

Qy	581	KGFPDEESIVNYTLNQAYQDNV--TVFASVIFQ---TRKDGSLPPHVKIR-----	627
Db	108	RGFPSEKDFEDY----IRYDNCSSSVLAADVFEHPFNHSKEPLPLAVKYHLRFYSYTRRNY	163
Qy	628	---QNSSFTEK-----TNEIRRAYWRPG-----PNTGGRFYFLYGFVWIQDMMERAI	672
Db	164	MWTQTGSFFLKETEGWHTTSLFPLFPNPGPRELTSPDGGEGPYIREGFLAVQHAVDRAIM	223
Qy	673	DTFVGHHDVVEPGSY-----VQMFPPYCYTRDDFLFVIEHMMPLCMVISWVYSVAMTIQH	726
Db	224	EYHA--DAATRQLFQRLTVTIKRFYPFFIADPFLVAIQYQLPLLLLLSFTYTALTIARA	281
Qy	727	IVAEKEHRLKEVMKTMGLNNAVHWAVFITGFVQLSISVTALTAIL-----KYGQVLMHS	781
Db	282	VVQEKERRLKEYMRMMGLSSWLHWSAWFLFLFFLFLLIAASFMTLLFCVKVKNVAVLSRS	341
Qy	782	HVVIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHD	841
Db	342	DPSLVLAFLLCFAISTISFSFMVSTFFSKANMAAFGGFLYFFTYIPYFFVAPR----YN	397
Qy	842	KITAFEKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVE-GDDFNLLLAVTML	900
Db	398	WMTLSQKLCSCLLSNVAMAMGAQLIGKFEAKGMIQWRDL-LSPVNVDDDFCFGQVLGML	456
Qy	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFFLQKSYWLGSGRTEAWESWPWARTPRLSV	960
Db	457	LLDSVLYGLVTWYMEAVFPQGFGVPQWPYFFIMPSYWCCKPRAVAGK-----	503
Qy	961	MEEDQACAMESRRFEETRGMEEPTHLPVVCVDKLTKVYK--DDKKLALNKLNLNLYEN	1018
Db	504	-EEEDSDPEKALRNEY---FEAEPEDLVAGIKIKHLSKVFRVGNKDRAAVRDLNLNLYEG	559
Qy	1019	QVVSFLGHNGAGKTTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLF	1078
Db	560	QITVLLGHNGAGKTTTTLSMLTGLFPPTSGRAYISGYEISQDMVQIRKSLGLCPQHDILF	619
Qy	1079	RLTVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAF	1138
Db	620	NLTVAEHLYFYAQLKGLSRQKCPEEVKQMLHIIGLEDKWNRSRSLSGGMRRKLSIGIAL	679
Qy	1139	VGGSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHG	1198
Db	680	IAGSKVLILDEPTSGMDAISRRAIWDLQRQKSDRTIVLTTHFMDEADLLGDRIAIMAKG	739
Qy	1199	KLKCCGSPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFI	1258
Db	740	ELQCCGSSSLFLKQKYGAGYHMTLVKEP-----HCNPEDISQLV	777

Qy	1259	RKHVASCLLVSDTSTELSYILPSEAAKKGAFERLFQHLERSLDALHLSSFGLMDTTLEEV	1318
Db	778	HHHVPNATLESSAGAELSFILPRESTHR--FEGLFAKLEKKQKELGIA SFGASITTMEEV	835
Qy	1319	FLKVSEEDQSLENSEADVKE SRKDVLPGAEGPASGEGHAGN LARCSELTQS QASLQSASS	1378
Db	836	FLRVGK----LVDSSMDIQAIQ---LPAIQ--YQHERRASDWAVDSNL-----	874
Qy	1379	VGSARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALS RVGQGSRKLDGGW-LKV	1437
Db	875	CGAMDPSDGIG-----ALIEEERTAV-----KLNTGLALHC	905
Qy	1438	RQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALS VP EIGDLPPLVLSPSQYHN	1497
Db	906	QQFWAMFLKKAAYS WREWKMVAAQVLVPLTCVT LALLAINYSSELFDDPMLRLTLGEY--	963
Qy	1498	YTQPRGNFI PYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCV-LKSPANGSLGPT	1556
Db	964	-----GRTVVPFSVPGTSQLGQQ	981
Qy	1557	LNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPD EDLQAWNVS L	1616
Db	982	LS-----	983
Qy	1617	PPTAGPEMWT SAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTDITG	1676
Db	984	-----EHLKDALQAE-----QEPREVLGDL-----	1004
Qy	1677	HNVSEYLLFTSDRFLHRYGAITFG--NVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNN	1734
Db	1005	---EEFLIFRA-----SVEGGGFNERCLVAASF-----RDVGERTVVNALFNN	1044
Qy	1735	KGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQG-TDV	1793
Db	1045	QAYHSPATALAVVDNLLF-----KLLCGPHA-SIVVS NFPQPRSA LQA AKDQFNEGRKGF	1098
Qy	1794	VIAIFIIVAMSFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDM LNYLVPAT	1853
Db	1099	DIALNLLFAMAFLASTFSILAVSERAVQAKHVQFVSGVHVASFWSALLWDLISFLIPSL	1158
Qy	1854	CCVIILFVFDLPAYTSPTNFP AVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINL	1913
Db	1159	LLLVVFKAFDVRAFTRDGHMADTLLLLLLYGWAIIPMYLMNFFFLGAATAYTRLTIFNI	1218
Qy	1914	FIGITATVATFLLQLFEHDKDLKV--VNSYLKSCFLIFPNYNLGHGLMEMAYNEY-----	1966
Db	1219	LSGI----ATFLMVTIMRIPAVKLEELSKTLDHVFLVLPNHCLGMAVSSF-YENYETRRY	1273
Qy	1967	-----INEYYAKIGQFDKMKSPFEWDI--VTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQ	2019
Db	1274	CTSSEVAAHYCKKYNIQYQENFYAWSAPGVGRFVASMAASGCAYLILLFLIETNLLQRLR	1333
Qy	2020	-----RMPVSTKPVEDDVDVASERQVR LRGDADNDM--VKIENLTKV	2059
Db	1334	GILCALRRRRTLTELYTRMPV---LPEDQDVADERTRILAPSPDSLHTPLIIKELSKV	1389

Qy 2060 YKSRKIGRILAVDRLCLGVRPGECFGLLGVNGAGKTSTFKMLTGDESTTGGEAFVNGHSV 2119  
 | : | : | | | | | | : | : | | | | | | | | : | : | | | | :  
 Db 1390 YEQRV--PLLAVDRLSLAVQKGECFGLLGFGAGKTTTFKMLTGEESLTSGDAFVGGHRI 1447

Qy 2120 LKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKY 2179  
 : : : | : | | | | | | : | : | | | : | : | : | :  
 Db 1448 SSDVGKVRQRIGYCPQFDALLDHMTGREMLVMYARLRGIPERHIGACVENTLRGLLLEPH 1507

Qy 2180 ADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLIILDIKTGRSV 2239  
 | : | | | | | | | | | | | | | | | | | | : | : | : : : : :  
 Db 1508 ANKLVRTYSGGNKRKLSTGIALIGEPAVIFLDEPSTGMDPVARRLLWDTVARARESGKAI 1567

Qy 2240 VLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVRTKS---SQSVKDV 2296  
 : : | | | | | | | | | | | | : : | | | | | : : : : : : : : : :  
 Db 1568 IITSHSMEECEALCTRLAIMVQGGQFKCLGSPQHLKSKFGSGYSLRAKVQSEGQOEAL EEF 1627

Qy 2297 VRFFNRNFP EAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDN 2356  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 1628 KAFVDLTFPGSVLEDEHQGMVHYHLPGRDLSWAKVFGILEKAKEKYGVDDYSVSQISLEQ 1687

Qy 2357 VFNFAKKQSDNLEQ 2371  
 | : : | | | : | :  
 Db 1688 VFSLFAHLQPPTAEE 1702

# RESULT 6

T33783

hypothetical protein Y39D8C.1 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 17-Mar-2000

C;Accession: T33783

R;Becker, M.; Graves, T.; Yoakum, M.

submitted to the EMBL Data Library, October 1998

A;Description: The sequence of *C. elegans* cosmid Y39D8C.

A;Reference number: Z21408

A;Accession: T33783

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1802 <BEC>

A;Cross-references: EMBL:AF101313; PIDN:AAC69223.1; GSPDB:GN00023; CESP:Y39D8C.1

A;Experimental source: strain Bristol N2; clone Y39D8C

C;Genetics:

A;Gene: CESP:Y39D8C.1

A;Map position: 5

A;Introns: 45/3; 114/1; 195/1; 230/3; 543/3; 794/1; 849/1; 1036/2; 1099/1;

1132/3; 1165/1; 1322/3; 1458/2; 1560/3; 1656/1

C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 16.0%; Score 2024; DB 2; Length 1802;

Best Local Similarity 28.6%; Pred. No. 2.8e-125;

Matches 539; Conservative 319; Mismatches 629; Indels 400; Gaps 48;

Qy 580 FKGFPEESIVNYTLNQAYQ--DNVTVFASVIFQTRKDGS LPPHVHYKIRQNSSFTEKTN 637  
 : | | | | : | : : | | : | : | : | : : : : | : :  
 Db 168 YKGFTTEGEMVSWMQGQFQSECDN-PLLAGIVF----DDSIKDLKNPDKRDFTYTIRLS 222

Qy 638 EIRR-----AY-W-----RPGPNTGGR-FYFLYGFVWIQDMMERAI 671  
 | : | | | | : | | : | : | : |  
 Db 223 NTHRRSRNAFGDNSYPWDTSVSFAVQYVSGPINPDDNDGSGPGYWQEGFMTVQRAVDVAI 282  
  
 Qy 672 IDTFVGHVDV-VEP--GSY-VQMFPYPCYTRDDFLFVIE---HMMPLCMVISWVYSVAMTI 724  
 : | | : | | | | | : | : | | : | : | : | :  
 Db 283 TEIITGEDAQLTPLLDYSYQVSRFPFPGYSTK----IIEIGAFFMPVIVIFSFMSTSVIYIV 338  
  
 Qy 725 QHIVAEKEHRLKEVMKTMGLNNAVHWWAWFITGFVQLSISVTALTAILKYQVLMHSHV 784  
 : : | | | | | | : | | : : | | | : : | : | :  
 Db 339 RAVVVEKEDRLKEYMRVMGLSQFINWVAHFIINYAKLTFVAVIVLTILMHF--VALKSDMT 396  
  
 Qy 785 I IWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIYFLSYVPYMYVAIREEVAHDKIT 844  
 : : : | | : | | : | : | : | : | : | : | :  
 Db 397 LMFVFLMIYAFDVVYFAFMISSEFMNSATSATLISVFWMLLYFWYAFFS-----SIDQTN 451  
  
 Qy 845 AFE---KCIASLMSTTAFGLGSKYFALYEAVGVGIQWHTFSQSPVEGDDFNLLLAVTMLM 901  
 : : | : | : | : | : | : | : | : | : | :  
 Db 452 PYPLGYRLINCINPDIALNYGLQLLAAYETQADGLKWGELFTPPSPDNNTFGHALIALI 511  
  
 Qy 902 VDAVVYGILTWYIEAVHPGMYGLP-RPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSV 960  
 | | : : | | | | | | | | | | : | : | : | | | |  
 Db 512 VDGIIIMIILTWYIEAVIPGGEGVPQKPWFVFL-PSYWF-----PNSGS 553  
  
 Qy 961 MEEDQACAMESRRFEETRGMEEEPTHLPLVVCVDKLTQVY-----KDDKKL 1006  
 | : : : : : : | : | : | : | : | : | : | :  
 Db 554 KTVDSDDQFQQIQYADHVKLEKEPTDLIPTINVVNLTQYGTSTFFKKLFDCKFGKSGEKR 613  
  
 Qy 1007 ALNKLSLNLYENQVVSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKN 1066  
 | : | : | : | | | | | | | | : | : | : | : | : | :  
 Db 614 AVSNLNLKMPGQCTVLLGHNGAGKSTTFSMLTGVASPSSGSAYVNDFDIRTSPLKIRRE 673  
  
 Qy 1067 LGMCPQHNVLFDRLTVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSIVQTL 1126  
 : | : | : | : | | : | | | | : : | : | : | : | :  
 Db 674 MGLCPQYNTLFGFMTVMEHLEFFAKLKERTWDP--EEAREILARLRIDFKADFMAGALSG 731  
  
 Qy 1127 GMKRKLSVAIAFVGGSRAILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEAD 1186  
 | | | | : | : | : | : | : | : | : | : | : | : | :  
 Db 732 GQKRKLSLAIALIGGSEVVMLDEPTSGMDPGARHETWTLIQREKERTILLTTHFMEEAD 791  
  
 Qy 1187 LLGDRIAIISHGKLKCCGSPLFLKGTYGDRYRLTLVKRPAEPGGPQEPGLASSPPGRAPL 1246  
 | | | | : : | : | : | : | : | : | : | : | : | :  
 Db 792 LLGDRIAIMAHGQLECCGSPMFLKQQYGDGYHLTIVY-----DTTSTP----- 834  
  
 Qy 1247 SSCSELQVSQFIRKHVASCLLVSDTSTELSYILPSEAACKGAERLQHLERSLDALHLS 1306  
 : : : | : : : | : | : | : | : | : | : | :  
 Db 835 ---DVSKTTDIIREYIPEAHVFSYIQEATYLL--SATHRPIFPKLFKELEDHQTCGKIT 889  
  
 Qy 1307 SFGIMDTTLEEVLKVSEEDQSLENSEADVKESSKDVLPGAEGPASGEGHAGNLARCSEL 1366  
 | | : | : | : | : | : | : | : | : | : | :  
 Db 890 SFGVSITTMEEVFLKVGHTADERYNYEHGIENDISEMI----- 927  
  
 Qy 1367 TQSQASLQSASSVGSARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQG 1426  
 : | : | : | : | : | : | : | : | : | :  
 Db 928 -----EKDDPILQDLRAQV----- 941  
  
 Qy 1427 SRKLDGGWLKVRQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLP 1486

```

      :: | |:: : || | |:: | :: : : : : | : :
Db      942 --RVTGFTLQMQHAKAMFYKRAIFFFRKWTQFLPQLVFPVAYLVLMVFTSQVLPSVKEQD 999

Qy      1487 PLVLSPSQYHNYTQPRGNFIPIYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLK 1546
      | : | |:: : : : | | | | |
Db      1000 PQTIS-----LAPFSDTKKAGH-----LVS-----DSGNYVTL---- 1027

Qy      1547 SPANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPD 1606
      | | : | |
Db      1028 -----LGGSQNLS----- 1035

Qy      1607 EDLQAWNVS LPPTAGPEMWT SAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVV 1666
      | | | | : |
Db      1036 -----SMVQGTVTQLGVT 1048

Qy      1667 TGDILTDITGHNVS EYLLFTSDRFLHRYGAITFG--NVLKSIPA--SFGTRAPPMVRKI 1722
      : | | | | : : : : | : | | | : | : | :
Db      1049 --QTVVDITS-NVEKFIMDQTNAM-----GSRTFGLHYALGFVPSMFNFSTVSVPSLK-- 1098

Qy      1723 AVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPM-NKTSAS 1781
      | : | | | : : : : | : | | | : | :
Db      1099 -----TFFNNFGLYTPALAITFTDSMIL-----SQKQKKQYSFTAVNHPLPPSTQDT 1145

Qy      1782 LSLDYLLQGTDVVIAIFIIVAMSFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANY 1841
      | | : | | : | : | : | : | : | : | : | :
Db      1146 LKNTNRSDGAAFLIAYGLIVSFAVCVAGYSQFLITERKKKSKHMQLLSGIRPWWFWLTAF 1205

Qy      1842 VWDMLNLYLPATCCVILFVFDLPAYTSPTNFPVAVLSL-FLLYGWSITPIMYPASFWEV 1900
      : | | : : | | : : : | | : : | | | | : | | : | |
Db      1206 IWDAAWFVIRILCFDAIFYIFNITAYTHDFGVMLILTLSELFLLYGWTALPFTYWFQFFFES 1265

Qy      1901 PSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVVNSYLKSCFL--IFPNYNLGHGL 1958
      : : : : : | : : | : : : : | | | : | | | :
Db      1266 APKGFMVMTMYHILTMIGSIAVPII-----SQTSSLDAGYLWSIIFAWLFPTYNISQIA 1320

Qy      1959 MEMAYNEYINEYYAKIGQFDKMKSPFEWDIVT-----RGL----- 1993
      | | : | : | : | : | :
Db      1321 TVTFQENENVRIACKKLDCTIPM-----FKAVTACCGTASERLYVDNVLFVGNRKGILVYV 1375

Qy      1994 VAMAVEGVVGFLLTIM---CQYNFLRRPQRMVSTKPVED-----DVDVASE 2037
      : : | : | : : | : | : | : | : | : | : | :
Db      1376 IFLAVQGFIYWIWFMRENDQFTKLFALIRCKRKNPIWDITDITDKVDERDVEDSDVIAE 1435

Qy      2038 RQRVLRGDADNDMVKI-ENLTQVYKSRKIGRILAVDRLCLGVRPGECFGLLGVNGAGKTS 2096
      : | | : | : | | | | | | : | : | | | | | |
Db      1436 KSVVQRLANNKKTALVSNNLVKQWY-----GNFNAVKGVNFEHVNSKDCFGLLGVNGAGKTS 1490

Qy      2097 TFKMLTGDESTTGGEAFVNGHSLVKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLR 2156
      | | : | | : | : | | | : : | | | | | : | : | :
Db      1491 TFQMLTGENSISSGDAYVNGWSVKNNWREAGANTGYCPQYDAIKEMSGETLYMFARIR 1550

Qy      2157 GISWKDEARVVKWALEKLELTQYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTG 2216
      | | | | : | : : : | : | | | | : | | : | | | : |
Db      1551 GIPEKDIPKKVNAVIHAIGIGMYASRQIKTYSGGNKRRLSLGIAIVGLPDVLLLDEPTSG 1610

Qy      2217 MDPKARRFLWNLILDLIKTRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNR 2276
      : | | | | : | : : | : | | | | : | | | | | : | : | :

```

Db 1611 VDPKARRIWNILNRLRDLGTALVLTSHSMDECEALCTELAIMVYGKFRCYGSCQHIKSR 1670

Qy 2277 FGDGYMITVRTKSSQSVKDVVRFFNRFPEAMLERHHTKVQYQLKSEHISLAQVFSKME 2336  
 :| || : :| |: : : : | :||| | : : : : | :||| |:|

Db 1671 YGSGYTLIRLKNRNDAEKTKSTIKQTFRGSVIKEEHVLQLNFDIPRDGDSWSRLFEEKLE 1730

Qy 2337 QVSGVLGIEDYSVSQTTLDNVFNFAK 2363  
 || | :|||:||||: ||: |::

Db 1731 TVSTSLNWDDYSLSQTTLEQVFIEFSR 1757

# RESULT 7

A84845

probable ABC transporter [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C;Accession: A84845

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell, C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams, M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver, G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: A84845

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1816 <STO>

A;Cross-references: GB:AE002093; NID:g6598351; PIDN:AAC02761.2; GSPDB:GN00139

C;Genetics:

A;Gene: At2g41700

A;Map position: 2

Query Match 15.5%; Score 1964.5; DB 2; Length 1816;  
 Best Local Similarity 26.9%; Pred. No. 2.6e-121;  
 Matches 580; Conservative 304; Mismatches 596; Indels 677; Gaps 63;

Qy 525 LHPEALNLSLDELPPALRQDNFSLPSGMALL-----QQLDTIDNAACGWIQFMS- 573  
 :|| |: | : : :||| :||| | :| | :|

Db 15 IHPAHSNIDKDTVVEVGKGNPSFPEVLKLLAEGDFLAFAPDDETNN----MIDILSL 70

Qy 574 -----KVSVDIFKGFPEESIVNYTLNQAYQ-----DNVTVFASVIFQTRKDGSLPP 620  
 :: ||| |: : | : | | : :||| : |

Db 71 KPPELRLVTKIFK---DDIELEYITSAHYGVCSEVRNCSNPKIKGAVVFHEQ-----GP 122

Qy 621 HV-HYKIRQNSSFTEKTNEIRRAYWRPGPN-----TGGRF-----YFLY 658  
 |: | ||| :: : || | | : |

Db 123 HLFDISIRLNHTWA-----FAGFPNVKSIMDTNGPYINDLEMGINTIPTMQYSFS 172

Qy 659 GFVWIQDMMERAIL-----DTFVGHDVVEPG-----SYVQMFPPYPC--YTR 697  
 ||: :| ::| || | : | : | :||| |:| ||

Db 173 GFLTLOQVVDSEFIIFASQQNNDLPLSHSNLSSALRFELPWTLFSPSVIRMVFPPTREYTD 232

Qy 698 DDFLFVIEHMMPLCMVISWVYSVAMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAVFITG 757  
|:| :: :| |

Db 233 DEFQSIVKSVMGL----- 245

Qy 758 FVQLSISVTALTAILKYGQVLMHSHVVIWFLAVYAVATIMFCFLVSVLYSKAKLASAC 817  
: || | :: : : : || |:| ::|| | |

Db 246 -----FLFKY-----SDKTLVFTYFFLFLGLSAIMLSFMISTFFTRAKTAVAV 287

Qy 818 GGIIYFLSYVPYMYVAIREEVAHDKITAF EKCIASLMSTTAFGLGSKYFALYEVAGVGIQ 877  
| : : :: || | : : | :||:| ||| ||| || ||| |||:

Db 288 GTLTFLGAFFPY-----TVNDESVMVLKVVASLLSPTAFALGSINFADYERAHVGLR 341

Qy 878 WHTFSQSPVEGDDFNLLAVTMLMVDVAVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYW 937  
| :: : : :|:::| | |:: | | : || | | :

Db 342 WSNIWRA---SSGVSFVCLLMMLLDSILYCALGLYLDKVLPRENGVRYPWNFIFSKYFG 398

Qy 938 LGSGRTEAWEWSWPWARTPRLSVMEEDQACAMESRRFEETRGMEEEPHTLPL----- 989  
: | : | | | :| :| :

Db 399 RKKNNLQ-----NRIPGFETD-----MFPADIEVNQGEFPDPVFESISLEMRQOE 443

Qy 990 ----VVCVDKLT KVYKDDKK--LALNKL SLNLYENQVVSFLGHNGAGKTTTMSILTGLFP 1043  
: | | ||| : |:| | | ||||:| |||||:|:|:| || |

Db 444 LDGRCIQVRNLHKVYASRRGNCCAVNSLQLTLYENQILSLLGHNAGKSTTISMLVGLLP 503

Qy 1044 PTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSRLKSMAQEEIRRE 1103  
||| | | |: | | ||||| |:|:|:| | | || :| : : :

Db 504 PTSGDALILGNSIITNMDEIRKELGVCPQHDLFP ELTVREHLEMF AVLKGVEEGSLKST 563

Qy 1104 MDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAFVGGSRAIILDEPTAGVDPYARRAIW 1163  
: | |:| |:| :|:| | |||||:| | :| :| | |||||:|:|:| | |

Db 564 VVDMAEEVGLSDKINTLVRALSGGMKRKLSLGIALIGNSKVIIILDEPTSGMDPYSMRLTW 623

Qy 1164 DLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPLFLKGT YGDGYRLTLVK 1223  
|| | | || |:| | |||:| |||| |:|:| | |||| :|| | | ||||

Db 624 QLIKKIKKGRIILLTTHSMDEAEELGDRIGIMANGSLKCCGSIFLKHHYGVGYTLTLVK 683

Qy 1224 RPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHVASCLLVSDTSTELSYILPSEA 1283  
:| | : : : :| | |:| :|:| |

Db 684 -----TSPTVSV-----AHIVHRHIPSATCVSEVGNEISFKLP--L 718

Qy 1284 AKKGAFERLFQHLERSL-----DALHLSSFGLMDTTLEEVFLKVSEEDQSLENSEADVKE 1338  
| || |:|:| : | : |:| | |||||:| :|:|

Db 719 ASLPCFENMFREIESCMKNSDSYDPIQSYGISVTTLEEVFLRVA-----GCNLDIED 771

Qy 1339 SRKDVLPGAEGPASGEGHAGNLCARSELTSQASLQSASSVGSARGDEGAGYTDVYGDYR 1398  
::|: : :| | : :| | : | | | :|

Db 772 KQEDIFVSPDTKSS-----LVC--IGSNQKSSMQPKLLASCNDGAGVIITSVAKAFR 821

Qy 1399 PL-----FDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQFHGLLVKRFH 1449  
: | : |:| :|| | | | :||

Db 822 LIVAAVWTLIGF-----ISIQCCGCSIISR-----SMFW---RHCKALFIKRAR 862

Qy 1450 CARRNSKALFSQILLPAFFV-----CVAMTVALSVP EI-----GDLP- 1486  
| |:| : | :|| | : :| | | : | :|

Db 863 SACRDRKTVAFQFIIPAVFLLFGLLFLQLKPHPDQKSITLTTAYFNPLLSGKGGGGPIPF 922

Qy 1487 ----PLVLSPSQY--HNYTQPRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVG 1540

Db	923	DLSVPIAKEVAQYIEGGWIQPLRN-----TSYKFPNPKE	956
Qy	1541	ATCVLKSPANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSD	1600
Db	957	ALADAIDAAGPTLGPTL-LSMSE--FLMSSFDQS--YQSSREGL-----SSHD	999
Qy	1601	SPASPEDELQAWNVS LPPTAGPEMWT SAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHP	1660
Db	1000	SCNHDPDGS�-----GYT-----	1011
Qy	1661	PQMRVVTGDILTDITGHNVS EYLLFTSDRFLHRYGAITFGNVLKSIPASFGTRAPPMVR	1720
Db	1012	-----	1011
Qy	1721	KIAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTS A	1780
Db	1012	-----VLHNGTCQHAGPIYINVMHAAILRL---ATGN-KNMTIQTRNHPLPPTKT	1057
Qy	1781	SLSLDYLLQGTDV---VIAIFIIVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYW	1837
Db	1058	Q-----RIQRHDLDAFSAAIIVNIAFSFIPASFVPIVKEREVKAKHQQLISGVSLSYW	1112
Qy	1838	LANYVWDMNLNVLVPATCCVILFVFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFW	1897
Db	1113	LSTYVWDFISFLFPSTFAILFYAFGLEQFIGIGRFLPTVLMLEYGLAIASSTYCI TFF	1172
Qy	1898	FE-----VPSSAYVF-----LIVINLFIGITATVATFLLQLFEHDKDLKV	1937
Db	1173	FTEHSMAQATSSYSVLLPISLFVFSFSSNVILMVHFFSGLILMVISFVMGLIPATAS---	1229
Qy	1938	VNSYLK-----SCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMK-SPFEW	1986
Db	1230	ANSYLKELILFRYALQNFFRLSPGFCFSDGLASLA-----LLRQGMKDKSSHGVFEW	1281
Qy	1987	DIVTRGLVAMAVEGVVGFLLTIMCQYNFL-----RRP	2018
Db	1282	NVTGASICYLGLEVRLEY-----CRYSMLLLSFFHGIDTKLSLIYTIGASRLTELIYDRV	1336
Qy	2019	QRMPVSTKP-----VEDDVDVASERQRVLRGDADNDMVKIENLTKVYKSRK-I	2065
Db	1337	YSTSFSTEP LLDSTGAISTDMEDDIDVQEERDRVISGLSDNTMLYLQNLRKVYPGDKHH	1396
Qy	2066	GRILAVDRCLGVRPGECFGLLG VNGAGKTSTFKMLTGDESTTGGEAFVNGHSVLKELLQ	2125
Db	1397	GPKVAVQSLTFSVQAGECFGLGTNGAGKTTTSLMLSGETPTSGTAFIFGKDIVASPKA	1456
Qy	2126	VQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELT KYADKPAG	2185
Db	1457	IRQHIGYCPQFDALFEYLTVKEHLELYARIKGVVDHRIDNVVTEKLVEFDLLKHSHKPSF	1516
Qy	2186	TYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLILD L- IKTGR-SVVLTS	2243
Db	1517	TLSGGNKRKLSVAIAMIGDPPIVILDEPSTGMDPVAKRFMWDVISRLSTRSGKTAVILT T	1576
Qy	2244	HSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGD-----	2279



```

Db      1577 HSMNEAQAALCTRIGIMVGGRLRCIGSPQHLKTRYGNHLELEVPFYNGVKPNEVSNVELEN 1636
Qy      2280 -----GYMITVVRTK-----SSQSVKDVVRFFNR- 2302
          :: | | : | : | : : |
Db      1637 FCQIIQQWLFNVPTQPRSLLDLEVCIGVSDSITPDASASEISLSPENVQRIAKFLGNE 1696
Qy      2303 -----NFPEA 2307
          : | | |
Db      1697 QRVSTLVPPLPEEDVRFDDQLSEQLFRDGGIPLPIFAEWWTKEKFSALDSFIQSSFPGA 1756
Qy      2308 MLKERHHTKVQYQLK--SEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFVNFA 2362
          | : : | | : | | | : | : | | : | | : | : |
Db      1757 TFKSCNGLSIKYQLPFGEGLSLADAFGHLERNRNLGIAEYSISQSTLETIFNHFA 1813

```

# RESULT 8

T47150

hypothetical protein DKFZp547P193.1 - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 02-Sep-2000

C;Accession: T47150

R;Bloeker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S.  
submitted to the Protein Sequence Database, March 2000

A;Reference number: Z24376

A;Accession: T47150

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-373 <AAA>

A;Cross-references: EMBL:AL162060

A;Experimental source: fetal brain; clone DKFZp547P193

C;Genetics:

A;Note: DKFZp547P193.1

C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette  
homology

```

Query Match          15.2%; Score 1920; DB 2; Length 373;
Best Local Similarity 100.0%; Pred. No. 1.3e-119;
Matches 373; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2064 KIGRILAVDRLCLGVRPGECFGLLVNGAGKTSTFKMLTGDESTTGGEAFVNGHSVLKEL 2123
          |||
Db      1 KIGRILAVDRLCLGVRPGECFGLLVNGAGKTSTFKMLTGDESTTGGEAFVNGHSVLKEL 60

Qy      2124 LQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKP 2183
          |||
Db      61 LQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKP 120

Qy      2184 AGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIKTRSVVLTS 2243
          |||
Db      121 AGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIKTRSVVLTS 180

Qy      2244 HSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVVRTKSSQSVKDVVRFFNRN 2303
          |||
Db      181 HSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVVRTKSSQSVKDVVRFFNRN 240

Qy      2304 FPEAMLERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFVNFAK 2363
          |||

```

Qy	655	YFLYGFWVIQ-----DMMERAIIDTFVGHDVVEPGSYVQMFYPYCYTRD	698
Db	19	:   ::  : ::    :           :	
Qy	699	DFLFVIEHMMPLCMVISWVYSVAMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAVFITGF	758
Db	79	:     ::  ::  : ::       ::     :     :   ::  :	
Qy	759	VQLSISVTALTAILKYGQVLMHSHVVIWLFLAVYAVATIMFCFLVSVLYSKAKLASACG	818
Db	137	::    :: :: ::   :: ::    ::  : ::  : ::	
Qy	819	GIIYFLSYVPYMYVAIREEVAHDKITA--FEKCIASLMSTTAFGLGSKY---FALYEVAG	873
Db	197	:: : :: : :  :::   :  : ::	
Qy	874	VGIQWHTFSQSPVEGDDFNLLAVTMLMVDVAVYGILTWYIEAVHPGMYGLPRPWYFPLQ	933

Db	250	WSDLWE--MNNPVLG--ISVELCMIMLVVDTAVFLILAWYISAVAPGDFGVRQPLWFFFT	305
Qy	934	KSYWLGSGRTEAWESWPWARTPRLSVMEEDQACAMESRRFEETRGMEEEPTHLPVVCV	993
Db	306	LKYWA-----PGLYKNRVEFVDDEHFDITPN-----SDSFDSEPTNL-----	342
Qy	994	DKLTKVYKDDKKLALNKLNLNLYENQVVSFLGHNGAGKTTTMSILTGLFPPTSGSATIYG	1053
Db	343	-----TLALDCLNLRLYEGQITGLLGHNGAGKTTTMSILCGLYAPSSGTAKIYQ	391
Qy	1054	HDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSRLKSMAQEEIRREMDKMIEDLEL	1113
Db	392	RDIRTDLRRVRDVLGICPQHNVLFSLTLVSEQRLRLFAALKGVDPSELTSQVDEILASVSL	451
Qy	1114	SNKRHSLVQTLSSGGMKRKLSVAIAFVGGSRAILDEPTAGVDPYARRAIWDLILKYKPGR	1173
Db	452	TEKANKLASTLSSGGMKRRLCIGIAFIGGSRFVILDEPTAGVDVTARKDIWKLLQRNKEGR	511
Qy	1174	TILLSTHHMDEADLLGDRIAIISHGKLKCCGSPLFLKGTYGDGYRLTLVKRPAEPGGPQE	1233
Db	512	TILLSTHHMDEADVLSDRIAILSQD-----FEKPDLLDGKRL-----	548
Qy	1234	PGLASSPPGRAPLSSCSELQVSQFIRKHVASCLLVSDTSTELSYILPSEAARKGAERLF	1293
Db	549	-----IF	550
Qy	1294	QHLESLDALHLSSFGLMDTTLEEVLKVSEEDQSLENSEADVKESSRKDVLPGAEGPASG	1353
Db	551	QH-----	552
Qy	1354	EGHAGNLARCSELTQSQASLQSSVSGSARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQ	1413
Db	553	-----	552
Qy	1414	EVEAEALSRVQGQSRKLDGGWLKVRQFHGLLVKRFHCARRNSKALFSQIILLPAFFVCVA-	1472
Db	553	-----FYALLVCRINYTLKSKRTFLFQVIIPFLFLLALAE	586
Qy	1473	----MTVALSVPEI-GDLPLVLSPSQYHNYTQPRGNFIPYANEERREYRLRLSPDASPQ	1527
Db	587	LFVLLQVSTARPDLMVSMPPPLETSIMGNHS----DF--YVNS-----WDTAENSTAN	634
Qy	1528	QLVSTFRLPSGVGATCVLKS PANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLS	1587
Db	635	DILHAMFSSPGTGPRCAKDVPND-----LLDTMRRELMFR-----N	670
Qy	1588	NFVPPPPSPAPSDSPASPEDLDQAWNVSLPPTAGPEMWTSAPSLPRLVREPVRCTCSAQG	1647
Db	671	RYGFGRNKPAPGVKDSVDNEYQCQNIQ-----GEFDYTEDIS-NATYNAPIYCGCEDFG	724
Qy	1648	TGFSCPSSVGHPQMRVVTGDILTITGHNVSEYLLFTSDRFRL-----HRYGAITF	1700
Db	725	WNCTLEDWKWNETNWLRLNTTDRIFDLTGRNLTQFRLIT--RFAQLANTTAPFFLGGFSL	782
Qy	1701	GNV-----LKSI-----PASFGTRAP-----	1716

Db 783 GHVNQRAQSQADIDTSKRGWLETIKDIAQSMRIINLNTTGIEPATPKVLDPFQAQNLTLNQ 842

Qy 1717 ---PMVRKIAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNH 1773  
 ::: : || :|::||| : | | |::|| :| || ||

Db 843 VVNDLLQNLQDRENVKVVFNKIVPGFPIASNILSNALLRQE--DYAIDPEDLGILTMNH 900

Qy 1774 PMNKTSASLSLDYLLQGTDVVIAIF----IIVAMSFVPASFVFLVAEKSTKAKHLQFVS 1829  
 ||||| | :|| :||| ::: :| :|| | |::| :| ||| |

Db 901 PMNKT-ISQTLQNALKFTQALAVFRITILLVLSMIPAGFTVYLVEDRICEALHLQLVG 959

Qy 1830 GCNPIIYWLANYVWDMNLNYPATCCVILFVFDLPAYTSPTNFPVLSLFLLYGWSITP 1889  
 | : ||: |::||: : | | : : | || ||

Db 960 GLRKVTYWVTSYLYDMVGGIHPRHHC-----NNAHLP-VLPCLRLRYRRRNRI 1005

Qy 1890 IMYPASFWFEVPSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSYLSKSCF--- 1946  
 : | | :|: | |:: |||

Db 1006 LRLPPS-----ILRARNVDDSLRL--CIPKSLFCGG 1034

Qy 1947 -LIFPNYNLGHGLMEMAYNEYINEYYAKIGQ-FDKMKSPFEWDIVTR----GLVAMAVEG 2000  
 | ||| : :: :|:: :| | |:: : |:: |

Db 1035 SLFCPNCN---WFLLRHSLCLDSYHARVAYGSEQMNRP---DMINQLPLPSLLAFDQMG 1088

Qy 2001 V-----VGFLLTIMC-----QYNFLRRPQR-----MPVSTKPVEDDVVASERQRLV 2042  
 : : : :| : : |:: :| | : || || ||||

Db 1089 IHIMCLFIHVIIATICLIFSQMDEFGFVRKRERNLTDAMMLREPSTCDEDDVVKERQRV- 1147

Qy 2043 RGDA-----DNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLGVNGAGKTS 2096  
 || : : ||| | :|| : | |||||::|||::

Db 1148 --DAIPMDSSDNHALIVRNLAAYNP----ELLAVKGISFAVEPGEÇFGLLGLNGAGKTT 1201

Qy 2097 TFKMLT-----GDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQL 2151  
 || ||| | | :| | : : || ||||| || :|: |::|:

Db 1202 TFAMLTAKIRPGHSGSIEMQNTTRINTGS-FSDVRNFQQ-LGYCPQFDALNMKLSTRENLF 1259

Qy 2152 YTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLD 2211  
 | |::||| : : | | | ||: : |||:|||| |::|: |::|||

Db 1260 YARIRGIVPTQIDSIIIDRLIALHLRPYANTQTSSLGGNRRKLSVAVALVSQPSLIFLD 1319

Qy 2212 EPTTGMDPKARRFLWNILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQ 2271  
 ||: |||| ::||| :| | |::| ||||| |||||::||| ||:|||| |

Db 1320 EPSAGMDPGSQQFLWKVIERLCKSGKAVVLTSHSMEECEALCTRIAIMDRGRIRCLGGKQ 1379

Qy 2272 HLKNRFGDGYMITVRTKSSQSVKDVVRFFNRNFP EAMLER-HHTKVQYQLKSEHISLAQ 2330  
 |||:::| | |::: : :|:: : | | : : : |::

Db 1380 HLKSKYGKGSMLTMKMGKDENAKEIAGIMRSKLGDSRVEAIHCSTIFIHIEQGTASVAR 1439

Qy 2331 VFSKMEQV 2338  
 | : ||

Db 1440 VLEIVNQV 1447

RESULT 10

C88925

protein F33E11.4 [imported] - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001

C;Accession: C88925

R;anonymous, The C. elegans Sequencing Consortium.  
 Science 282, 2012-2018, 1998  
 A;Title: Genome sequence of the nematode C. elegans: a platform for  
 investigating biology.  
 A;Reference number: A75000; MUID:99069613; PMID:9851916  
 A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and  
 www\_sanger.ac.uk/Projects/C\_elegans/ for a list of authors  
 A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103,  
 1999; and Science 285, 1493, 1999  
 A;Accession: C88925  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-1317 <STO>  
 A;Cross-references: GB:chr\_V; PIDN:AAC17542.1; PID:g3158495; GSPDB:GN00023;  
 CESP:F33E11.4  
 C;Genetics:  
 A;Gene: F33E11.4  
 A;Map position: 5

Query Match 13.3%; Score 1688; DB 2; Length 1317;  
 Best Local Similarity 28.1%; Pred. No. 3.5e-103;  
 Matches 441; Conservative 241; Mismatches 480; Indels 406; Gaps 31;

```

Qy      858 AFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTMLMVDVVYGILTWYIEAV 917
      |  | :  | | |  |::|  |  ::  | :  |::| ::  | | | | | |
Db      49 ALNYGLQLLAAYETQADGLKWGELFTPPSPDNNLTFGHALIALIVDGIIMIILTWYIEAV 108

Qy      918 HPGMYGLP-RPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSVMEEDQACAMESRRFEE 976
      ||  |::| :||:| |  || |  |  | :  :  :: :
Db      109 IPGGEVGPQKPWFVFL-PSYWF-----PNSGSKTVDSSDQFQQIQYAD 150

Qy      977 TRGMEEEPTHLPLVVCVDKLTQVY-----KDDKKLALNKLNLNLYENQVVS 1022
      :||| |  : |  || |  |  | : | |:: |:: | : |
Db      151 HVKLEKEPTDLIPTINVVNLTKTYGTSFFKKLFDCKFGKSGEKRAVSNNLNMYPGQCTV 210

Qy      1023 FLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRITV 1082
      | | | | | | | : | | : | | : | | | : | | | : | | : | | : | |
Db      211 LLGHNGAGKSTTFSMLTGVASPSGSAAYVNDFDIRTSLPKIRREMGLCPQYNTLFGFMTV 270

Qy      1083 EEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVGG 1142
      || | |:::| |  :  | :: | : |  :  || | | | | : | | : | |
Db      271 MEHLEFFAKLKERTWDP--EEAREILARLRIDFKADFMAGALSGGQKRKLSLAIALIGGS 328

Qy      1143 RAILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKC 1202
      :: | | | | : | | |  | | : |  | | | : | | : | | | | | | : | : |
Db      329 EVVMLDEPTSGMDPGARHETWTLIQREKERRTILLTTHFMEEADLLGDRIAIMAHGQLEC 388

Qy      1203 CGSPLFLKGTG DYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHV 1262
      | | | : | | | | | | : |  | : |  : :  | | : :
Db      389 CGSPMFLKQQYGDGYHLTIVY-----DTTSTP-----DVSKTTDIIREYI 428

Qy      1263 ASCLLVSDTSTELSYILPSEAAKKGA FERLFQHLERSLDALHLSSFGLMDTTLEEVFLKV 1322
      : |  | : | : |  | : | : | : |  : | | : | | | | |
Db      429 PEAHVFSYIGQEATYLL--SATHRPIFPKLFKELEDHQTCGITSFGVSITTMEEVFLKV 486

Qy      1323 SEEDQSLENSEADV KESRKDVLPGAEGPASGEGHAGNLARCSELTQSQASLQSASSVGSA 1382
      | |  ::  :::

```

Db	487	GHTADERYNYEHGIENDISEMI-----	508
Qy	1383	RGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQFHG	1442
		:   :     :   : :     : :	
Db	509	-----EKDDPILQDLRAQV-----RVTGFTLQMQHAKA	536
Qy	1443	LLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVP EIGDLPPIVLSPSQYHNYTQPR	1502
		:      :   : : : : :   :	
Db	537	MFYKRAIFFFRKWTFQLPQLVFPVAYLVLMVFTSQVLPSPVKEQDPQTIS-----	585
Qy	1503	GNFIPIYANEERREYRLRLSPDASPQQVLSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1562
		: : : : :             :	
Db	586	--LAPFSDTKKAGH-----LVS---DSGNYVTL-----LGGSQNLS--	616
Qy	1563	ESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDELQAWNVS LPPTAGP	1622
Db	617	-----	616
Qy	1623	EMWTSAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHP PQMRVVTGDILTITGHNVSEY	1682
		:   :        :	
Db	617	-----SMVQGTVTQLGVT--QTVVDITS-NVEKF	642
Qy	1683	LLFTSDFRLHRYGAITFG--NVLKSIPASFGTRAPP MVRKIAVRRAAQVFYNNGKYHSM	1740
		:: ::   :       :   :       :	
Db	643	IMDQTNAM----GSRTFGLHYALGFVPSMF-----NFSTVSV	675
Qy	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVINHPMNKTSASLSLDYLLQGT DVVIAIFII	1800
		: :	
Db	676	PS-----LK	679
Qy	1801	VAMSFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLAN YVWDMNLNYPATCCVIILF	1860
		:: :   :    :   :    :   :    :    :    :   :	
Db	680	ISFAVCVAGYSQFLITERKKKSKHMQLLSGIRPWMFWLTAFIWDAAWF VIRILCFDAIFY	739
Qy	1861	VFDLPAYTSPTNFP AVL SL-FLLYGWSITPIMPASFWEVPSSAYVFLIVINLF IGITA	1919
		:   : :     :   :         :       :   : : : :   :	
Db	740	IFNITAYTHDFGVMLILTLSFLLYGW TALPFTYWFQFFFESAPKGFMVMTYHILTMIG	799
Qy	1920	TVATFLLQLFEHDKDLKV VNSYLKSCFL--IFPNYNLGHGLMEMAYNEYINEYYAKIGQF	1977
		::   : : : :     :     :   :   :	
Db	800	SIAPPII-----SQTSSL DAGYLWSIIFAWLFPTYNISQ IATVTFQENENVRIACKKL DCT	854
Qy	1978	DKMKSPFEWDIVT-----RGL---VAMAVEGVVGFLLTIM---	2009
		:    : : :   : :   :	
Db	855	IPM-----FKAVTACCGTASERLYVDNVLFVGN RKGILVYVIFLAVQGFIYWIWVFMREN	909
Qy	2010	CQYNFLRRPQRMPVSTKPVED-----DVDVASERQ RVLRGDADNDMVKI-EN	2055
		:     :   :     :   :     :   :	
Db	910	DQFTKLFALIRC RKADNPIWDITDTDKVDERDVEDSDVIAEKS SVVQRLANNNTALVS NN	969
Qy	2056	LTKVYKSRKIGRILA VDRLCLGVRPGECFGL LGVNGAGKTSTFKMLTGDESTTGGEAFVN	2115
		:   :                 :   :   :   :	
Db	970	LVKWY-----GNFNAVKGVNFHVNSKD C FGL LGVNGAGKTSTFQMLTGENS ISSGDAYVN	1024
Qy	2116	GHSV LKELIQVQQS LGYCPCQDALFDELTAREHL QLYTRLRGISWKDEARVVKWALEKLE	2175
		: :           :   :   :   :     :   :	
Db	1025	GWSVKNNWREAGANTGYCPOYDAI IKEMSGETLYMFARIRGIPEKDIPKKVNAV IHAIG	1084

Qy 2176 LTKYADKPAGTYSGGNKRRLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIKT 2235  
 : || : |||||:| ||:| | : ||||:|:||||| :|:: |  
 Db 1085 IGMYASRQIKTYSGGNKRRLSLGIAIVGLPDVLLLDDEPTSGVDPKARRIWNILNRLRDL 1144

Qy 2236 GRSVVLTSMSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVRTKSSQSVKD 2295  
 | ::|||:|:||||| |||| | : ||-|| ||:|:| | | : : | : :  
 Db 1145 GTALVLTSHSMDECEALCTELAIMVYGKFRCYGSCQHIKSRYGSGYTLLIRLKNRNDAEK 1204

Qy 2296 VVRFFNRNFP EAMLERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLD 2355  
 : | ::|| | :: : : : | ::| | : | | | :||:||||:  
 Db 1205 TKSTIKQTFRGSVIKEEHVLQLNFDIPRDGDSWSRLF EKLETVSTSLNWD DYSLSQT TLE 1264

Qy 2356 NVFVNFAK 2363  
 ||: |::  
 Db 1265 QVFIEFSR 1272

# RESULT 11

S60124

transport protein homolog C48B4.4 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 13-Jan-1996 #sequence\_revision 12-Apr-1996 #text\_change 02-Feb-2001

C;Accession: S60124; S40724; S40725

R;Kershaw, J.

submitted to the EMBL Data Library, November 1995

A;Reference number: S60124

A;Accession: S60124

A;Molecule type: DNA

A;Residues: 1-1767 <KER>

A;Cross-references: EMBL:Z29117; NID:g439247; PID:g1066912

C;Genetics:

A;Map position: III

A;Introns: 47/1; 112/3; 161/2; 220/2; 543/3; 574/3; 903/2; 1056/1; 1115/3;  
 1178/3; 1265/2; 1331/3; 1416/3; 1703/3

C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette  
 homology

C;Keywords: ATP; duplication; nucleotide binding; P-loop; transmembrane protein

F;628-818/Domain: ATP-binding cassette homology <ABC1>

F;645-652/Region: nucleotide-binding motif A (P-loop)

F;764-769/Region: nucleotide-binding motif B

F;1457-1642/Domain: ATP-binding cassette homology <ABC2>

F;1474-1481/Region: nucleotide-binding motif A (P-loop)

F;1586-1591/Region: nucleotide-binding motif B

Query Match 12.0%; Score 1524; DB 2; Length 1767;

Best Local Similarity 25.4%; Pred. No. 4.9e-92;

Matches 530; Conservative 332; Mismatches 638; Indels 584; Gaps 75;

Qy 447 LGLLVHLM-----TSNPKILY--APAGSEVDRVILKAN-----ETFAFVGNVT 487  
 || ||:|: ||:|: | | : | |::| : || :  
 Db 103 LGPLVYLVVKNADHTSSPENIYDNFQVKGTVEDVFLESNFIKPIYKRWCLRSDDVVVGYS 162

Qy 488 HYAQVWLNISAEIRSFLEQGRLLQQLRWLQQYVAELRLHPEALNLSLDELPPALRQDNFS 547  
 | : ::| | | | | :|:| : ' | ::| || |  
 Db 163 KDAAAKRTVDDLMMKKFAE--RFQS-----AKLKL SVKN-ESSEEQLLT VLRND--- 207

Qy 548 LPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQDNVTVFAS 607  
| : : | : : | : | | : | |  
Db 208 -----LPMLNETFCAINSYAAGV---VF---DEVDTVNKKLN----- 238

Qy 608 VIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRF-----YFLYG 659  
| : | : | : | : | : | : | :  
Db 239 -----YRILLGKT-PEETWHLTETSYPYGPSSGRYSRIPSSPPYWTSA 281

Qy 660 FVWIQDMMERAIIDTFVGHVDVEPGS-----YVQMFPYPCYTR-----DDFLFVIEHM 707  
| : | : | : : | : | : : | | | | : | :  
Db 282 FLTFQHAIESSFLSS-----VQSGAPDLPITLRGLPEPRYKTSSVSFAFIDFFPFI---- 331

Qy 708 MPLCMVISWVYSVAMTIQHI---VAEKEHRLKEVMKTMGLNN---AVHWWAVFITGFVQ 760  
| : : : | : | : | : | : | : | : | : | : | :  
Db 332 -----WAFVTFINVIHITREIAAENHAVKPYLTAMGLSTFMFYAAHVMAFLKFFVI 383

Qy 761 LSISVTALTAILKYQVLMHSHVVIWLFLAVYAVATIMFCFLSVLY---SKAKLASA 816  
| : | : : : : : : : | : : : | : : | : | :  
Db 384 FLCSIPLTFVMEF---VSPAALIVTVLM--YGLGAVIFGAFVASFFNNTNSAIKAILV 437

Qy 817 CGGIYFLSYVPYMYVAIREEVAHDKITAFKEC-IASLMSTTAFGLGSKYFALYEVAGVG 875  
| : : | : | : | : : | : | : | : : | : :  
Db 438 AWGAMIGISY-----KLRPEL--DQISS---CFLYGLNINGAFALAVEAISDYMRRERE 486

Qy 876 IQ-WHTFSQSPVEGDDEFNLLAVTMLMVDVAVYGILTWYIEAVHPGMYGLPRPWYFPLQK 934  
: : | : | : | : | : | : | :  
Db 487 LNLTNMFENDSSLH---FSLGWALVMMIVDIL----- 514

Qy 935 SYWLGS-----RTEAWESWPWART-----PRLSVMEEDQACAMESRRFEETR-- 979  
| : | : | : | : | : | : | : | : | : | :  
Db 515 --WSIGALVVDHIRTSA-DFS---LRTLDFEAPEDDENQTDGVTAQNTRINEQVRNRV 568

Qy 980 ----MEEEPHTLPLV-----VCVDKLTKVYKDDKKLAL 1008  
| | | : : | : | : | : : | : :  
Db 569 RRSDMEMNPMASSTSLNPPNADSDSLLEGSTEADGARDTARADIIVRNLVKIWSTTGERAV 628

Qy 1009 NKLSLNLYENQVVSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLG 1068  
: | | | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 629 DGLSLRAVRGQCSILLGHNGAGKSTTFSSIAGIIRPTNGRITICGYDVGNEPGETRRHIG 688

Qy 1069 MCPQHNVLFDRLTVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSVQTLSGGM 1128  
| | | : | : | : | | | : : : : : | : | : | : | : | : | :  
Db 689 MCPQYNPLYDQLTVSEHLKLVYGLKGAREKDFKQDMKRLSDVKLDFKENEKAVNLSGGM 748

Qy 1129 KRKLSVAIAFVGGSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLL 1188  
| | | | : | : | : : | : | : | : | : | : | : | : | : | : | : | :  
Db 749 KRKLCVCMALIGDSEVLLDEPTAGMDPGARQDVQKLVEREKANRTILLTTHYMDEAERL 808

Qy 1189 GDRIAIISHGKLKCCGSPLFLKGTYG DYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSS 1248  
| : : | : | : | : : | : | : | : | : | : | : | :  
Db 809 GDWVFIMSHGKLVASGTNQYLKQKFGTGYLLTVV---LDHNGDK----- 849

Qy 1249 CSELQVSQFIRKHAVASCLLVSDTST-----ELSYILPSEAAKKGAFERLFQ 1294  
| | : : : | : | : : | : | : | : | : | : | : | :  
Db 850 -----RK---MAVILTDVCTHYVKEAERGEMHGQOIEIILPE--ARKKEFVPLFQ 894

Qy 1295 HLE-----RSLDALHLSSFGLMDTTLEEVFLKVSEEDQSLNSEA 1334



Db	895	ALEAIQDRNYRSNVFDNMPNTLKSQLATLEMRSFGLSLNTLEQVFITIGDK---VDKAIA	951
Qy	1335	DVKESRKDVLPGAEGPASGEGHAGNLCARSELTSQASLQASASSVGSARGDEGAGYTDVY	1394
Db	952	SRQNSR-----ISHNSRNASPSLKPAGYDTQSSTKSA-----	984
Qy	1395	GDYRPLFDNPQDPDNVSLQEVEAEALSRVQGGSRKLDGGWLKVRQFHGLLVKRFHCARRN	1454
Db	985	DSYQKLMD-----SQARGPEKSGVAKM-----VAQFISIMRKKFLYSRRN	1024
Qy	1455	SKALFSQILLPAFFVCVAMTVALSVPEIGDLPLVL-SPSQYHNYTQPRGNFIPYANEER	1513
Db	1025	WAQLFTQVLPIIILLGL-----VGSLTTLKSNNTDQFRSLT-----	1060
Qy	1514	REYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSESRLIAARFFD	1573
Db	1061	-----PSGIEPSKVWVRFENGTI-----	1078
Qy	1574	SMCLESFTQGLPLSNFVPPPPSPAPSDSPASPEDLQAWNVS LPPTAGPEMWTSAPSLPR	1633
Db	1079	-----PEE-----AANFEK	1087
Qy	1634	LVREPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILTDITGHNVS EYLLFTSDRFLH	1693
Db	1088	ILRKS-----GGF-----EVLNNTKNPL-----	1106
Qy	1694	RYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYSMP TYLNSINNAILR	1753
Db	1107	-----PNITKSL---IGEMPPATIGMTMSDNLEALFNMRYYHVLPTLISMIN---R	1152
Qy	1754	ANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLL--QGTDVVIAIFIIVAMS FVPASFV	1811
Db	1153	ARLTGTVD AEISSGVFL-----YSKSTSNSNLLPSQLIDVLLAPMLILIFAMVTSTFV	1205
Qy	1812	VFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVII LFVFDLPAYTSPT	1871
Db	1206	MFLIEERTCQFAHQQLTGISPITFYASLIYDGILY---SLICLIFLFMF-LAFHWMYD	1261
Qy	1872	NFPAVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITATVATFLL-QLFE	1930
Db	1262	HLAIVILFWFLYFFSSVPFIYAVSFLFQSPSKANVLLI IWQVVISGAALLAVFLIFMIFN	1321
Qy	1931	HDKDLK--VVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFEWDI	1988
Db	1322	IDEWLKSILVNIFM----FLLPSYAFGSAIIT-----INTY----GMILPSEELMNWDH	1367
Qy	1989	VTRGLVAMAVEGVVGFLLTIMCQYNFLRR--PQMPV-----STKPVEDDV----DVAS	2036
Db	1368	CGKNAWLMGTFGVCSFALFVLLQFKFVRRFLSQVWTVRRSSHNNVQPMMGDLPVCEVSE	1427
Qy	2037	ERQRVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLG VNGAGKTS	2096
Db	1428	ERERVHRVNSQNSALVIKDLTKTF-----GRFTAVNELCLAVDQKECFGLLG VNGAGKTT	1482
Qy	2097	TFKMLTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLR	2156

Db	1483	TFNILTQGSEASSGEAMIGGRDV-TELI----	SIGYCPQFDALMLDLTGRESLEILAQM	1537
Qy	2157	GI-SWKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTT		2215
Db	1538	GFENYKAKAELI---LECVGMIAHADKLVRFYSGGQKRKISVGVALLAPTQMIILDEPTA		1594
Qy	2216	GMDPKARRFLWNLIILDIKTGRS-VVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLK		2274
Db	1595	GIDPKARREVWELLLLWCREHSNSALMLTSHSMDECEALCSRIAVLNRGSLIAIGSSQELK		1654
Qy	2275	NRFGDGYMITVRTKSSQSVKDVVRFNRNPFPEAMLKERHHTK----	VQYQL-KSEHISLAQ	2330
Db	1655	SLYGNNTMTLSLYEPNQDMVVLVQTRLPNLSVLKTTSTNKTLNLKWQIPKEKEDCWSA		1714
Qy	2331	VFSKMEQVSGVLGIEDYSVSQTTLDNVFVNFAKKQSDNLEQQET		2374
Db	1715	KFEMVOALAKDLGVKDFILAOSSLEETFLRLAGLDEDOLDTHST		1758

Query Match 12.0%; Score 1522.5; DB 2; Length 1758;  
Best Local Similarity 25.5%; Pred. No. 6.1e-92;  
Matches 528; Conservative 331; Mismatches 644; Indels 569; Gaps 74;

Qy	548	LPSGMALLQQIDTIDNAACGWIQFMSKVSVDIFKGFPEDEESIVNYTLNQAYQDNVTVFAS	607
Db	208	-----LPMLNETFCAINSYAAGV---VF---DEVDVTNKKLN-----	238
Qy	608	VIFQTRKDGSLPPHVHYKIRQNSSTFEKTNEIRRAYWRPGPNTGGRF-----YFLYG	659
Db	239	-----YRILLGKT-PEETWHLTETSYPYGPSSGRYSRIPSSPPYWTSA	281
Qy	660	FVWIQDMMERAIIDTFVGHDVVEPGS-----YVQMFPPYCYTR-----DDFLFVIEHM	707
Db	282	FLTFQHAIESSFLSS-----VQSGAPDLPITLRGLPEPRYKTSSVSAFIDFFPFI----	331
Qy	708	MPLCMVISWVYSVAMTIQHI---VAEKEHRLKEVMKTMGLNN---AVHWWAVFITGFVQ	760
Db	332	-----WAFVTFINVIHITREIAAENHAVKPYLTAMGLSTFMFYAAHVMAFLKFFVI	383
Qy	761	LSISVTALTAILKYGQVLMHSHVVIWLFLAVYAVATIMFCFLVSVLY---SKAKLASA	816
Db	384	FLCSIIPLTFVMEF---VSPAALIVTVLM--YGLGAVIFGAFVASFFNNTNSAIKAILV	437
Qy	817	CGGIIYFLSYVPYMYVAIREEVAHDKITAFEKC-IASLMSTTAFGLGSKYFALYEVAGVG	875
Db	438	AWGAMIGISY-----KLRPEL--DQISS---CFLYGLNINGAFALAVEAISDYMRRERE	486
Qy	876	IQ-WHTFSQSPVEGDDFNLLLAVTMLMVDVAVYGILTWYIEAVHPGMYGLPRPWYFPLQK	934
Db	487	LNLTNMFNDSSLH---FSLGWALVMMIVDIL-----	514
Qy	935	SYWLGSG-----RTEAWEWSWPWART-----PRLSVMEEDQACAMESRRFEETRGM	981
Db	515	--WMSIGALVVDHIRTSA-DFS---LRTLDFEAPEDDENQTDGVTQAQNTTRINEQMNPMA	568
Qy	982	EEPTHLPLV-----VCVDKLTKVYKDDKKLALNKLNLNLYENQ	1019
Db	569	STSLNPPNADSDSLLEGSTEADGARDTARADIIVRNLVKIWSTTGERAVDGLSLRAVRGQ	628
Qy	1020	VVSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR	1079
Db	629	CSILLGHNGAGKSTTFSSIAGIIRPTNGRITICGYDVGNEPGETRRHIGMCPQYNPLYDQ	688
Qy	1080	LTVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFV	1139
Db	689	LTVSEHLKLVLVYGLKGAREKDFKQDMKRLLSDVKLDFKENEKAVNLSGGMKRKLVCMALI	748
Qy	1140	GGSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGK	1199
Db	749	GDSEVLLDEPTAGMDPGARQDVQKLVEREKANRTILLTTHYMDEAERLGDWVFIMSHGK	808
Qy	1200	LKCCGSPLFLKGTGYDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIR	1259
Db	809	LVASGTNQYLKQKFGTGYLLTVV---LDHNGDK-----R	839
Qy	1260	KHVASCLLVSDTST-----ELSYILPSEAAKKGAFERLFQHLE-----	1297
Db	840	K---MAVILTDVCTHYVKEAERGEMHQQIEIILPE--ARKKEFVPLFQALEAIQDRNYR	894

Qy 1298 -----RSLDALHLSSFGLMDTTLEEVFLKVSEEDQSLENSEADVKE SRKDVLP 1345  
| | : ||| |||:|: : : : : | : ||

Db 895 SNVFDNMPNTLKSQLATLEMRSFGLSLNTLEQVFITIGDK---VDKAIASRQNSR----- 946

Qy 1346 GAEGPASGEGHAGNLCARSELTSQASLQASASSVGSARGDEGAGYTDVYGDYRPLFDNPQ 1405  
: | || : | : || || | : | |

Db 947 -----ISHNSRNASPSLKPAGYDTQSSTKSA-----DSYQKLMD--- 981

Qy 1406 DPDNVSLQEVEAEALSRVQGSRKLDGGWLKVRQFHGLLVKRFHCARRNSKALFSQILLP 1465  
::| | | : | || :| :|| | :| :| :

Db 982 -----SQARGPEKSGVAKM-----VAQFISIMRKKFLYSRRNWAQLFTQVLIP 1024

Qy 1466 AFFVCVAMTVALSVP EIGDLPLVLSPSQYHNYTQPRGNFIPYANEERREYRLRLSPDAS 1525  
: : :| | | : | : |

Db 1025 IILLGL-----VGSLTTL-----KSN-----N 1041

Qy 1526 PQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLP 1585  
| ||| : : | || :

Db 1042 TDQFSVRSLTSPSGIEPSKVVRWFENGTI----- 1069

Qy 1586 LSNFVPPPPSPAPSDSPASPDDELQAWNVS LPPTAGPEMWTSA PSLPRLVREPVRCTCSA 1645  
| : | | : : :| :

Db 1070 -----PEE-----AANFEKILRKS----- 1083

Qy 1646 QGTGFSCPSSVGGHPPQMRVVTGDILTDITGHNVS EYLLFTSDRFR LHRYGAITFGNVLK 1705  
|| : : | | | : |

Db 1084 -----GGF-----EVLNNTKNPL-----PNITK 1102

Qy 1706 SIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAA 1765  
| : | | : : :| :| :| :| :| :| :| :| :

Db 1103 SL---IGEMPPATIGMTMNSDNLEALFNMRYYHVLPTLISMIN---RARLTGTVD AEIS 1155

Qy 1766 YGITVTNHPMNKTSASLSLDYLL--QGTDVVIAIFIIVAMSFVPASFVFLVAEKSTKAK 1823  
| : : | | | || | ||:| :| :| :| :| :| :| :

Db 1156 SGVFL-----YSKSTSNSNLLPSQLIDVLLAPMLILIFAMVTSTFVMFLIEERTCQFA 1208

Qy 1824 HLQFVSGCNPIIYWLANYVWDMNLNVLVPATCCVILFVFDLPAYTSPTNFPAVLSLFLLY 1883  
| ||:| :| :| :| :| :| :| :| :| :| :| :| :| :

Db 1209 HQQFLTGISPITFYASLIYDGILY---SLICLIFLFMF-LAFHWMYDHLAIVILFWFLY 1264

Qy 1884 GWSITPIMYPASFWEVPSSAYVFLIVINLFIGITATVATFLL-QLFEHDKDLK--VVNS 1940  
:| | :| ||| :| | ||| :| | :| :| :| :

Db 1265 FFSSVPFIYAVSFLFQSPSKANVLLIIWQVVISGAALLAVFLIFMIFNIDEWLKSILVNI 1324

Qy 1941 YLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEG 2000  
: : :| :| :| :| :| :| :| :| :| :| :

Db 1325 FM---FLLPSYAFGSAIIT-----INTY---GMILPSEELMNWDHCGKNAWLMGTFG 1370

Qy 2001 VVGFLLTIMCQYNFLRR--PQRMPV-----STKPVEDDV---DVASERQRVLRGDADN 2048  
| | | :| :| :| | | : :| :| :| :| :| :

Db 1371 VCSFALFVLLQFKFVRRFLSQVWTVRRSSHNNVQPMGDLPCESVSEERERVHRVNSQN 1430

Qy 2049 DMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLVNGAGKTSTFKMLTGDESTT 2108  
: | :| | : | | | :| | | | | | | | | :| :| :

Db 1431 SALVIKDLTKTF-----GRFTAVNELCLAVDQKECFGLLVNGAGKTTTFNILTQSFAS 1485

Qy 2109 GGEAFVNGHSVLKELLQVQQLGYCPQCDALFDELTAREHLQLYTRLRGI-SWKDEARVV 2167

```

      ||| : | | ||: |:||||| ||| :|| || |:: :: | :|| :| ::
Db      1486 SGEAMIGGRDV-TELI----SIGYCPQFDALMLDLTGRESLEILAQMhGFENYKAKAELI 1540

Qy      2168 KWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWN 2227
      || : : :||| |||| |||:| :||: | |||| |:||||| :|
Db      1541 ---LECVGMIAHADKLVRFYSGGQKRKISVGVALAPTQMIILDEPTAGIDPKARREVWE 1597

Qy      2228 LILDLIKTGRS-VVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVR 2286
      |:| : | ::|||||:|||||:|:|:| | | :|| | ||: |:| | :|:
Db      1598 LLLWCREHSNSALMLTSHSMDECEALCSRIAVLNRGSLIAIGSSQELKSlyGNNYTMtLS 1657

Qy      2287 TKSSQSVKDVVRFFNRNFPEAMlKERHHTK---VQYQL-KSEHISLAQVFSKMEQVSGVL 2342
      ||: | :|| | | :||:| : :||:| : : | : : :|
Db      1658 LYEPNQDMVVQLVQTRLPNsvLkTTSTNkTLNLKWQIPKEKEDCWSAKFEMVQALAKDL 1717

Qy      2343 GIEDYSVSQTTLDNVFVNFAKKQSDNLEQQET 2374
      |:|: :|:|:| :| : | | :| |
Db      1718 GVKDFILAQSSLEETFLRLAGLDEDQLDTHST 1749

```

# RESULT 13

T42749

ATP-binding cassette transport protein homolog - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 21-Jul-2000

C;Accession: T42749

R;Wu, Y.C.; Horvitz, H.R.

Cell 93, 951-960, 1998

A;Title: The *C. elegans* cell corpse engulfment gene *ced-7* encodes a protein similar to ABC transporters.

A;Reference number: Z22259; MUID:98297348; PMID:9635425

A;Accession: T42749

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-1704 <WUY>

A;Cross-references: EMBL:AF049142; NID:g3172340; PIDN:AAC24116.1; PID:g3172341

C;Genetics:

A;Note: *ced-7*

C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 12.0%; Score 1515; DB 2; Length 1704;

Best Local Similarity 25.4%; Pred. No. 1.8e-91;

Matches 529; Conservative 332; Mismatches 642; Indels 582; Gaps 75;

```

Qy      447 LGLLVHLM-----TSNPKILY--APAGSEVDRVILKAN-----ETFAFVGNVT 487
      || ||:|: ||:|: | | :| |::| : || :
Db      36 LGPLVYLvvKNADHTSSPENIYDNFQVKGTVEDVFLESNFIKPIYKRwCLRSdVVVGyTS 95

Qy      488 HYAQVWLNISAEIRSFLEQGRlQQHLRWLQQYVAELRLHPEALNLSLDELPPALRQDNFS 547
      | : :| | | | | | :|:| : | :| || |
Db      96 KDAAAKRTVDDLMKKFAE--RFQS-----AKLKLSVKN-ESSEEQLLTvLRND--- 140

Qy      548 LPsGMALLQQLDTIDNAACGWIQFMskVSVDIFKGFPDEESIVNYTLNQAYQDNVTVFAS 607
      | : : | : : | :| | | :| ||
Db      141 -----LPMLNETFCAINSyaAGV---VF----DEVdVTNKKLN----- 171

```

Qy 608 VIFQTRKDGSLPPHVHYKIRQSSFTEKTNEIRRAYWRPGPNTGGRF-----YFLYG 659  
 | : | : | : | : | : | : | :  
 Db 172 -----YRILLGKT-PEETWHLTETSYPYGPSSGRYSRIPSSPPYWTSA 214

Qy 660 FVWIQDMMERAIIDTFVGHDVVEPGS-----YVQMFPYPCYTR-----DDFLFVIEHM 707  
 | : | : | : | : | : | : | : | : | : | :  
 Db 215 FLTFQHAIESSFLSS-----VQSGAPDLPITLRGLPEPRYKTSSVSFAIDFFPFI---- 264

Qy 708 MPLCMVISWVYSVAMTIQHI---VAEKEHRLKEVMKTMGLNN---AVHWVAFITGFVQ 760  
 | : : : | : | : | : | : | : | : | :  
 Db 265 -----WAFVTFINVIHITREIAAENHAVKPYLTAMGLSTFMFYAAHVMAFLKFFVI 316

Qy 761 LSI SVTALTAILKYGQVLMHSHVVIWFLAVYAVATIMFCFLVSVLY-----SKAKLASA 816  
 | : | : : : : : : : | : : : | : : | :  
 Db 317 FLC SIPLTFVMEF----VSPAALIVTVLM--YGLGAVIFGAFVASFFNNTNSAIKAILV 370

Qy 817 CGGIYFLSYVPYMYVAIREEVAHDKITAFEKC- IASLMSTTAFGLGSKYFALYE VAGVG 875  
 | : : | : | : | : | : | : | : | : | :  
 Db 371 AWGAMIGISY-----KLRPEL--DQISS---CFLYGLNINGAFALAVEAISDYMRRE 419

Qy 876 IQ-WHTFSQSPVEGDDFNLLAVTMLMVDVAVYGILTWYIEAVHPGMYGLPRPWYFPLQK 934  
 : : | : | : | : | : | : | : | :  
 Db 420 LNLTNMFENDSSLH---FSLGWALVMMIVDIL----- 447

Qy 935 SYWL GSG-----RTEAWESWPPWART-----PRLSVMEEDQACAMESRRFEETR-- 979  
 | : | | | : : | | | : | : | : | : | :  
 Db 448 --WMSIGALVVDHIRTSA-DFS---LRTLDFEAPEDDENQTDGVT AQNTRINEQVRNRV 501

Qy 980 ----MEEEP THLPLV-----VCVDKLT KVYKDDKKL 1006  
 | : : : : : : : : : : : : : : : :  
 Db 502 RRS DMEIQMNPMASTSLNPPNADSDSLLEGSTEADGARDTARADIIVRN LVKIWSTTGER 561

Qy 1007 ALNKL SLNLYENQVVSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKN 1066  
 | : : | | | | | | | | | : | : | : | : | : | : | :  
 Db 562 AVDGLSLRAVRGQCSILLGHNGAGKSTTFSSIAGIIRPTNGRITICGYDVGNEPGETRRH 621

Qy 1067 LGMCPQHNVLFDRLTVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTL SG 1126  
 : | | | : | : | : | | | | : : : : | : : | : | : | :  
 Db 622 IGMCPQYNPLYDQLTVSEHLKL VYGLKGAREKDFKQDMKRLLSDVKLDFKENEKAVNL SG 681

Qy 1127 GMKRKLSVAIAFVGGSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEAD 1186  
 | | | | | | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 682 GMKRKLCVCMALIGDSEVLLDEPTAGMDPGARQDVQKLVEREKANRTILLTTHYMDEAE 741

Qy 1187 LLGDRIAIISHGKLKCCGSPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPL 1246  
 | | : | : | | | : : | : | : | : | : | : | :  
 Db 742 RLGDWVFIMSHGKLVASGTNQYLKQKFGTGYLLTVV---LDHNGDK----- 784

Qy 1247 SSCSELQVSQFIRKHHVASCLLVSDTST-----ELSYILPSEAAKKGA FERL 1292  
 | | : : : | : | : : | : | : | : | :  
 Db 785 -----RK---MAVILTDVCTHYVKEAERGEMHGQQIEIILPE--ARKKEFVPL 827

Qy 1293 FQHLE-----RSLDALHLSSFGLMDTTLEEVFLKVSEEDQSLENS 1332  
 | | | | | : | : | : | : | : | : | : : : : :  
 Db 828 FQALEAIQDRNYRSNVFDNMPNTLKSQ LATLEMRSFGLSLNTLEQVFITIGDK---VDKA 884

Qy 1333 EADV KESRKDVLPGAEGPASGEGHAGNLARCELTQSQASLQSASSVGSARGDEGAGYTD 1392

Db	885	IASRQNSR-----ISHNSRNASEPSLKPAGYDTQSSTKSA-----	919
Qy	1393	VYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQFHGLLVKRFHCAR	1452
Db	920	--DSYQKLMD-----SQARGPEKSGVAKM-----VAQFISIMRKKFLYSR	957
Qy	1453	RNSKALFSQILLPAFFVCVAMTVALSVP EIGDLPLVLSPSQYHNYTQPRGNFIPYANEE	1512
Db	958	RNWAQLFTQVLIPILLGL-----VGSLTTL-----KSN-----	986
Qy	1513	RREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFF	1572
Db	987	-----NTDQFSVRSRLTPSGIEPSKVVRWFENGTI-----	1015
Qy	1573	DSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPEDLQAWNVSIPPTAGPEMWTSAAPSLP	1632
Db	1016	-----PEE-----AANFE	1023
Qy	1633	RLVREPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILTDITGHNVS EYLLFTSDRFRL	1692
Db	1024	KILRKS-----GGF-----EVLNYNTKNPL-----	1043
Qy	1693	HRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSMP TYLNSLNNAIL	1752
Db	1044	-----PNITKSL---IGEMPPATIGMTMNSDNLEALENMRYYHVLPTLISMIN----	1088
Qy	1753	RANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLL--QGTDVVIAIFIIIVAMSFVPASF	1810
Db	1089	RARLTGTVD AEISSGVFL-----YKSTSNSNLLPSQLIDVLLAPMLILIFAMVTSTF	1141
Qy	1811	VVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVIIIFVFDLPAYTSP	1870
Db	1142	VMFLIEERTCQFAHQFLTGISPITFYASSLIYDGILY---SLICLI FLFMF-LAFHWMY	1197
Qy	1871	TNFPVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGITATVATFLL-QLF	1929
Db	1198	DHLAIVILFWFLYFFSSVPFIYAVSFLQSPSKANVLLIIWQVVISGAALLAVFLIFMIF	1257
Qy	1930	EHDKDLK--VVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFEWD	1987
Db	1258	NIDEWLKSILVNIFM----FLLPSYAFGSAIIT-----INTY----GMILPSEELMNWD	1303
Qy	1988	IVTRGLVAMAVEGVVGFLLTIMCQYNFLRR--PQRM PV-----STKPVEDDV----DVA	2035
Db	1304	HCGKNAWLMGTFGVCSFALFVLLQFKFVRRLS QVWTVRRSSHNNVQPMMGDLPVCESVS	1363
Qy	2036	SERQVRLRGDADNDMVKIENLT KVYKSRKIGRILAVDRLCLGVRPGECFGLLG VNGAGKT	2095
Db	1364	EERERVHRVNSQNSALVIKDLTKTF-----GRFTAVNELCLAVDQKECFGLLG VNGAGKT	1418
Qy	2096	STFKMLTGDESTTGGEAFVNGH SVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRL	2155
Db	1419	TTFNILTGQSFASSGEAMIGRDV-TELI----SIGYCPQFDALMLDLTGRESLEILAQM	1473
Qy	2156	RGI-SWKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPT	2214





Db	158	FREIESCMKNSDSYPGIQSYGISVTTLEEVFLRVA-----GCNLDIEDKQEDIFVSP	210
Qy	1348	EGPASGEGHAGNLARCSELTQSQASLQSASSVGSARGDEGAGYTDVYGDYRPL-----	1400
Db	211	DTKSS-----LVC--IGSNQKSSMQPKLLASCNDGAGVIITSVAKAFRLIVAAVWTL	260
Qy	1401	--FDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQFHGLLVKRFHCARRNSKAL	1458
Db	261	IGF-----ISIQCCGCSIISR-----SMFW---RHCKALFIKRARSACRDRKTV	301
Qy	1459	FSQILLPAFFV-----CVAMTVALSVEI-----GDLP-----PLVLS	1491
Db	302	AFQFIIPAVFLLFGLLFLQLKPHPDQKSITLTTAYFNPLLSGKGGGGPIPFDLSPVPIAKE	361
Qy	1492	PSQY--HNYTQPRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPA	1549
Db	362	VAQYIEGGWIQPLRN-----TSYKFNPKEALADAIDAA	395
Qy	1550	NGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPEDL	1609
Db	396	GPTLGPTL-LSMSE--FLMSSFDQS--YQSSREGL-----SSHDSCNHPDGSL	438
Qy	1610	QAWNVS LPPTAGPEMWTSA PSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGD	1669
Db	439	-----GYT-----	441
Qy	1670	ILTDITGHN VSEYLLFTSDRFR L HRYGAITFGNV LKSIPASFGTRAPPMVRKIAVRRAAQ	1729
Db	442	-----	441
Qy	1730	VFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQ	1789
Db	442	VLHNGTCQHAGPIYINVMHAAILRL---ATGN-KNMTIQTRNHPLPPTKTQ-----RIQ	491
Qy	1790	GTDV---VIAIFIIVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDM L	1846
Db	492	RHDLDAFSAATIVNIAFSFIPASFAVPIVKEREVKAKHQQLISGVS VLSYWLSTYVWDFI	551
Qy	1847	NYLVPATCCVILFVEDLPAYTSPTNFPAVLSLFLLYGWSITPIMYPASEWFE-----	1899
Db	552	SFLFPSTFAILFYAFGLEQFIGIGRFLPTVLM LLEYGLAIASSTYCLTFFFTTEHSMAQA	611
Qy	1900	-----VPSSAYVF-----LIVINLFIGITATVATFLLQLFEHDKDLKVNSYLK---	1943
Db	612	TSSYSVLLPISL FVFSFSSNVILMVHFFSGLILMVISFVMGLIPATAS---ANSYLKELI	668
Qy	1944	-----SCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMK-SPFEWDIVTRGLVA	1995
Db	669	LFRYALQNFFRLSPGFCFSDGLASLA-----LLRQGMKDKSSHGVFEWNVGTGASICY	720
Qy	1996	MAVEGVVGFLLTIMCQYNFL-----RRPQRMVSTKP	2027
Db	721	LGLEVRLEY-----CRYSM LLSFFH GIDTKLSLIYTIGASRLTELIYDRVYSTSFSTEP	775
Qy	2028	-----VEDDVDVASERQRVLRGDADNDMVKIENLTKVYKSRK-IGRILAVDRL	2074

Db	776	LLKDSTGAISTDMEDDIDVQEERDRVISGLSDNTMLYLQNLRKVYPGDKHHGPKVAVQSL	835
Qy	2075	CLGVRPGEFCFGLLGVNAGKTSTFKMLTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCP	2134
Db	836	TFSVQAGECFGFLGTNGAGKTTTSLMSLGEETPTSGTAFIFGKDIVASPKAIRQHIGYCP	895
Qy	2135	QCDALFDELTAREHLQLYTRLRGISWKDEARVVWALEKLELTKYADKPAGTYSGGNKRK	2194
Db	896	QFDALFEYLTVKEHLELYARIKGVVDHRIDNVVTEKLVEFDLLKHSKPSFTLSGGNKRK	955
Qy	2195	LSTAIALIGYPAFIFLDEPTTGMDPKARREFLWNILIDL-IKTGR-SVVLTSHSMEECEAL	2252
Db	956	LSVAIAMIGDPPIVILDEPSTGMDPVAKRFMWVDVISRLSTRSGKTAVILTTHSMNEAQAL	1015
Qy	2253	CTRLAIMVNGRLRCLGSIQHLKNRFGD-----GYM	2282
Db	1016	CTRIGIMVGGRLRCIGSPQHLKTRYGNHLELEVPPFYNGVKPNEVSNVELENFCIIQQWL	1075
Qy	2283	ITVRTK-----SSQSVKDVVRFFNR-----	2302
Db	1076	FNVPQTQPRSLLDGDLVCIGVSDSITPDASASEISLSPENVQRIAKFLGNEQRVSTLVPP	1135
Qy	2303	-----NFPEAMLKERHHTK	2316
Db	1136	LPEEDVRFDQLSEQLFRDGGIPLPIFAEWWTKEKFSALDSFIQSSFPGATFKSCNGLS	1195
Qy	2317	VQYQLK--SEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFNFA	2362
Db	1196	IKYQLPFGEGGLSLADAFGHLERNRNLGIAEYSISQSTLETIFNHFA	1243

Query Match 9.5%; Score 1202.5; DB 2; Length 1564;  
Best Local Similarity 24.4%; Pred. No. 9.5e-71;  
Matches 444; Conservative 280; Mismatches 544; Indels 549; Gaps 61;

Qy	685	SYVQMFPPCYTRDDF-----LFVIEHMMPLCMVISWVYSVAMTIQHIVAKEKH	733
		: :   : :   : :   : :   : :   : :   : :   : :	
Db	175	SIENFDIKLYSITNFEESGFGQNFGLLFAVCMIMP-----VISVA---RALVVEKS-	223
Qy	734	RLKEVMKTMGLNNAVHWVAVFITGFVQLSISVTAL-----	768
		:   :   :   : :   :   : : :	
Db	224	SVKPYLTTIGLPLWMFYLEHFLFGVIKNTFLITLLSTLYIFSMDCNCPYVLAGIFMYTCH	283
Qy	769	-----TAILKYGQVLMHSHVVIWLFLAVYAVATIMFCF-----LVSVLYSKAKL	813
		:   :   : : :   :   : :   :   : :   :	
Db	284	CVSFSILCTSILPFGKRIVEG-MVIIWITLIIAMHLSLEFEFDWLFVWVPLLNPNYSLKLF	342
Qy	814	ASACGGIIYFLSYVPYMYVAIREEVAHDKITAFEKCIASLMSTTAFGLGSKYFALYEVAG	873
		: :   : :   : :   : :   : :   : :   : :   : :	
Db	343	VDAT-----FLASGPN-----GTPTSALF----	361
Qy	874	VGIQWHTFSQSPVEGDDFNLLAVTMLMVDVVY--GILTWYIEAVHPGMYGLPRPWYFP	931
		:   :   : :   : :   : :   : :   : :   : :	
Db	362	--SSKKKTLQSAAVY---FGIMISCTVVMLVAAIFMEKLYTFVGHAI-----	404
Qy	932	LQKSYWLGSGRTEAWESWPWARTPRLSVME-----EDQACAMESRRFEETRGMEEEPTH	986
		:   : :   : :   : :   : :   : :   : :	
Db	405	--KRFRWILG-----FSKGKRSKIEERGDGVEDRSTILQCKETVEGRGSAIADIE	452
Qy	987	LPLVVCVDKLTKVYKDDKKLALNKLNLNLYENQVVSFLGHNGAGKTTTMSILTGLFPPTS	1046
		:   :   :   :   :   :   :   :   :   :   :   :	
Db	453	L-----SGLVKVYQNGEK-AVNGLSLRAIRGQVSILLGHNGCGKSTTFGMITGMHQATE	505
Qy	1047	GSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSRLKSMA-QEEIRREMD	1105
		:   :   :   :   :   :   :   :   :   :   :   :	
Db	506	GKVMIGGIDANANRAEARELIGYCPQYNPIYDELTVWEHLRLVNALKGRSGGSDFKMDAE	565
Qy	1106	KMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAFVGGSRAILDEPTAGVDPYARRAIWDL	1165
		: : :   : :   : :   : :   : :   : :   : :   : :   : :	
Db	566	SLLKQIELTDKRNLTAKNLSGGMKRKLVCMMAMIGGSRVILLDEPTAGMDPSARIDVQNM	625
Qy	1166	ILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLCCGSPLFLKGTYGDGYRLTLV-KR	1224
		:   :   :   :   :   :   :   :   :   :   :   :	
Db	626	LALVKADRTILLTTHYMDEAEKLGDWIFVMSHGKMAASGSKHYLKQYGGGMLTLVFKS	685
Qy	1225	PAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHVASCLLVSDTSTELSYI-----	1278
		:   :	
Db	686	VHDPMPRPK-----SYETAYDVCKTVCSALVKDERGQMIEISILETE	728
Qy	1279	---LPS-----EAAKKGAFERLFQHLERSLDALHLSSFGLMDTTLEEVFLKVSEE	1325
		: :   : :   : :   : :   : :   : :   : :   : :	
Db	729	KSRLPTLLKILESVMEEEDYNNPEFQALEPDIQEKCRTELELATIGVSMSSLEQVFIKIGDE	788
Qy	1326	-DQSELENSEADVKESSRKDVLPGAEGPASGEGHAGNLARCELTQSQASLQSASSVGSARG	1384
		: :   : :   : :   : :   : :   : :   : :   : :	
Db	789	CDDIMNGTGVDKKTERQE-----KFSTLVQYKI-----	816
Qy	1385	DEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQSGSRKLDGGWLKVRQFHGLL	1444
		: :	
Db	817	-----QQPK-----QGFSKL-----MMVWALL	834
Qy	1445	VKRFHCARRNSKALFSQILLPA---FFVCVAMTVALSVP EIGDLPLVLSPSQYHNYTQ	1500

Db	835	QKRAYYLNRNPVQITLQIILPLLTWLWFAVPFLRLEPKPPKLSDIES--FDPSQYPHST-	891
Qy	1501	PRGNFIPYANEERREYRLRLSPDASPQQVLSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Db	892	-----VLLQLEN-----	898
Qy	1561	SGESRLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDELQAWNVSLEPPTA	1620
Db	899	ENDDRL--ANYLNS-----FSNF-----	914
Qy	1621	GPEMWTSAAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTDTIGHNVS	1680
Db	915	-----EVVFKT-----LGFIVKVNKKGDSKFYKISQGD-----KNA	946
Qy	1681	EYLLFTSDRFLHRYGAITFGNVLSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Db	947	-----ILMNIIASAMYLDRDPSVTK-----	965
Qy	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFII	1800
Db	966	---LPHVTSRVIWMNDPKIK-----YEGLASFFLFEN-----IFFL	998
Qy	1801	VAMSFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWMLNYLVPATCCVILF	1860
Db	999	LVLAGIFIQSTVYLIEEKICKFAHQYLTGLSTIAYWGVFLWDFL-----LF	1046
Qy	1861	VFDLPAYTSPTNFPVAVLSLFLLYG--WSITPIMYPASFVFEVPSAYVFLIVINLFIGIT	1918
Db	1047	TFFL-LYT----IGFLISFGVLQGHIHEIVVIFYGLLFYF--APLVYLTSA LIN-----T	1094
Qy	1919	ATVATFLLQLF-----EHDKDLKVNSYLKSCFLIF-PNYNLGHGLMEMAY	1963
Db	1095	PTRGNFLLYMFCCIPWLAYSIVSELHNFPPPIQKYSDEIEYGFRIFNPSIGFLAGLMKIA-	1153
Qy	1964	NEYINEYYAKIG---QFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIM-----	2009
Db	1154	--ALN--YPKSGLDKHFEHLTNLWTYEGIFFELMFLEFFGGI---FLTILLGCATLKPFRR	1206
Qy	2010	-CQYNFLRRPQ-RMPVSTKPVEDDVVASERQVRVLRGDADNDMVKIENLTKVYKSRKIGR	2067
Db	1207	ACFRGTRRRSQPREKRKYKGIESC KAVKEEQVLVQEVDKNETVLVIDGLVKDF-----GK	1261
Qy	2068	ILAVDRCLGVRPGECFGLLG VNGAGKTSTFKMLTGDESTTGGEAFVNGH SVLKELLQVQ	2127
Db	1262	FRAVNDLSISVGHEECFGMLGANGAGKTTTFDIITGLTMPTGGSATIDGH DITETI----	1317
Qy	2128	QSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTY	2187
Db	1318	-HIGYCPQFDAMLQQISCRQTLRIMAKLQG--YPNVKEVVELVLDCVGM SDFGYKLVKNC	1374
Qy	2188	SGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIKTGR-SVVLTS HSM	2246
Db	1375	SGGQKRKISVGIALMSRATCIILDEPTAGIDPRARREIWDIIHEMREQAKCSIVLTSHSM	1434
Qy	2247	EECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVRTKSSQSVKDVVRFFN RNFPE	2306

Db 1435 EECEALCTRIGILRKGEMIALGTSQSLKSQYGNTYMMTLILNSLEDLESVCVIVSEEMPD 1494  
 Qy 2307 AMLKERHH---TKVQYQL-KSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFNFA 2362  
 |:|| | : ::| ||: :: ::::| :: :|| ::| :|:: |:   
 Db 1495 AVLKTPESSLTTSIVWELPKSKSDKWSEKYNQVEVLAKKANAKDYMLTQASLEDTFIRL- 1553  
 Qy 2363 KKQSDNLEQQETEPPIA 2379  
 : :| | ||   
 Db 1554 -----ITTEEEEEASA 1564

Search completed: September 1, 2004, 10:58:17  
 Job time : 94 secs

OM protein - protein search, using sw model

Run on: September 1, 2004, 10:57:18 ; Search time 217 Seconds  
(without alignments)  
3531.784 Million cell updates/sec

Title: US-10-088-467-2  
Perfect score: 12668  
Sequence: 1 MGFLHQQLQLLWKNVTLKRR.....GLISFEEERAQLSFNTDTLC 2436

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1297172 seqs, 314612898 residues

Total number of hits satisfying chosen parameters: 1297172

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	Query	
No.	Score Match Length DB ID	Description

1	12660	99.9	2436	16	US-10-380-727-2	Sequence 2, Appli
2	12656	99.9	2436	9	US-09-795-693-8	Sequence 8, Appli
3	12656	99.9	2436	14	US-10-156-239-8	Sequence 8, Appli
4	12656	99.9	2436	14	US-10-199-485-8	Sequence 8, Appli
5	10249	80.9	2001	13	US-10-072-621-8	Sequence 8, Appli
6	9237	72.9	1771	15	US-10-297-022-18	Sequence 18, Appl
7	7117	56.2	1472	14	US-10-340-097-119	Sequence 119, App
8	7117	56.2	1472	14	US-10-336-215-119	Sequence 119, App
9	7117	56.2	1472	14	US-10-336-219-119	Sequence 119, App
10	4240.5	33.5	2261	12	US-10-617-334-1	Sequence 1, Appli
11	4240.5	33.5	2261	15	US-10-452-510-1	Sequence 1, Appli
12	4240.5	33.5	2261	16	US-10-745-377-5	Sequence 5, Appli
13	4240.5	33.5	2261	16	US-10-744-465-1	Sequence 1, Appli
14	4237.5	33.5	2261	14	US-10-313-641-9	Sequence 9, Appli
15	4237.5	33.5	2261	14	US-10-313-641-10	Sequence 10, Appl
16	4237.5	33.5	2261	15	US-10-428-551-9	Sequence 9, Appli
17	4237.5	33.5	2261	15	US-10-428-551-10	Sequence 10, Appl
18	4234.5	33.4	2261	10	US-09-984-827-127	Sequence 127, App
19	4233.5	33.4	2261	9	US-09-995-542-11	Sequence 11, Appl
20	4233.5	33.4	2261	9	US-09-846-456-11	Sequence 11, Appl
21	4233.5	33.4	2261	10	US-09-984-827-2	Sequence 2, Appli
22	4233.5	33.4	2261	10	US-09-984-827-131	Sequence 131, App
23	4233.5	33.4	2261	10	US-09-984-827-134	Sequence 134, App
24	4232.5	33.4	2261	10	US-09-984-827-128	Sequence 128, App
25	4232.5	33.4	2261	10	US-09-984-827-129	Sequence 129, App
26	4232.5	33.4	2261	10	US-09-984-827-133	Sequence 133, App
27	4232.5	33.4	2261	10	US-09-984-827-136	Sequence 136, App
28	4232.5	33.4	2263	12	US-10-276-774-2326	Sequence 2326, Ap
29	4230.5	33.4	2261	10	US-09-984-827-135	Sequence 135, App
30	4229.5	33.4	2261	10	US-09-984-827-132	Sequence 132, App
31	4223.5	33.3	2261	10	US-09-984-827-130	Sequence 130, App
32	4192.5	33.1	2261	14	US-10-340-097-118	Sequence 118, App
33	4192.5	33.1	2261	14	US-10-336-215-118	Sequence 118, App
34	4192.5	33.1	2261	14	US-10-336-219-118	Sequence 118, App
35	4141	32.7	2201	12	US-10-170-385-293	Sequence 293, App
36	4141	32.7	2201	15	US-10-331-496A-29	Sequence 29, Appl
37	4141	32.7	2201	16	US-10-429-160-4	Sequence 4, Appli
38	4103	32.4	2201	9	US-09-995-542-9	Sequence 9, Appli
39	3951	31.2	2310	9	US-09-995-542-10	Sequence 10, Appl
40	3875.5	30.6	2273	12	US-10-182-006-6	Sequence 6, Appli
41	3857.5	30.5	2273	9	US-09-995-542-12	Sequence 12, Appl
42	3834.5	30.3	2273	14	US-10-340-097-3	Sequence 3, Appli
43	3834.5	30.3	2273	14	US-10-336-215-3	Sequence 3, Appli
44	3834.5	30.3	2273	14	US-10-336-219-3	Sequence 3, Appli
45	3834.5	30.3	2273	15	US-10-295-027-1279	Sequence 1279, Ap

#### ALIGNMENTS

##### RESULT 1

US-10-380-727-2

; Sequence 2, Application US/10380727

; Publication No. US20040024183A1

; GENERAL INFORMATION:

; APPLICANT: INCYTE GENOMICS, INC.; LEE, Ernestine A.;

```

; APPLICANT: YUE, Henry; LAL, Preeti G.;
; APPLICANT: CHAWLA, Narinder K.; BAUGHN, Mariah R.;
; APPLICANT: WARREN, Bridget A.; LEE, Sally;
; APPLICANT: SANJANWALA, Madhu S.; YAO, Monique G.;
; APPLICANT: RAMKUMAR, Jayalaxmi; THORNTON, Michael;
; APPLICANT: GANDHI, Ameena R.; POLICKY, Jennifer L.;
; APPLICANT: ELLIOTT, Vicki S.; ARVIZU, Chandra;
; APPLICANT: RAUMANN, Brigitte E.; BRUNS, Christopher M.;
; APPLICANT: NAINA, Amir; HAFALIA, April J.A.;
; APPLICANT: NGUYEN, Danniel B.; XU, Yuming;
; APPLICANT: LU, Dyung Aina M.; ISON, Craig H.;
; APPLICANT: GRIFFIN, Jennifer A.; REDDY, Roopa M.;
; APPLICANT: BURFORD, Neil
; TITLE OF INVENTION: TRANSPORTERS AND ION CHANNELS
; FILE REFERENCE: PI-0217 USN
; CURRENT APPLICATION NUMBER: US/10/380,727
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: PCT/US01/28938
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: US 60/241,700
; PRIOR FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US 60/240,540
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/239,057
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: US 60/236,882
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: US 60/234,842
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US 60/232,685
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 2436
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20040024183A1 7078207CD1
US-10-380-727-2

```

```

Query Match          99.9%; Score 12660; DB 16; Length 2436;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2435; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 MGFLHQLQLLLWKNVTLKRRSPWVLAFEIFIPLVLFILLGLRQKKPTISVKEVPFYTAA 60
        |||
Db      1 MGFLHQLQLLLWKNVTLKRRSPWVLAFEIFIPLVLFILLGLRQKKPTISVKEVSFYTAA 60

Qy     61 PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLDRVVEEGNLFDPARPSLGSE 120
        |||
Db     61 PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLDRVVEEGNLFDPARPSLGSE 120

Qy    121 LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSPNSTAQAL 180
        |||
Db    121 LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSPNSTAQAL 180

```



Qy	181	LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPALLEQLTC	240
Db	181	LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPALLEQLTC	240
Qy	241	TPGSSELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAELRNQLDVAKVSQQ	300
Db	241	TPGSSELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAELRNQLDVAKVSQQ	300
Qy	301	GLDAPNGSDSSPQAPPPrRLQALLGDLLDAQVLQDVVDLSALALLLPQGACTGRTPGPP	360
Db	301	GLDAPNGSDSSPQAPPPrRLQALLGDLLDAQVLQDVVDLSALALLLPQGACTGRTPGPP	360
Qy	361	ASGAGGAANGTGAGAVMGPNATAEEGAPSAAAALATPDTLQGQCASFVQLWAGLQPILCGN	420
Db	361	ASGAGGAANGTGAGAVMGPNATAEEGAPSAAAALATPDTLQGQCASFVQLWAGLQPILCGN	420
Qy	421	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF	480
Db	421	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF	480
Qy	481	AFVGNVTHYAQVWLNISAEIRSFLEQGRLQOHLRWLQQYVAELRLHPEALNLSDELPPA	540
Db	481	AFVGNVTHYAQVWLNISAEIRSFLEQGRLQOHLRWLQQYVAELRLHPEALNLSDELPPA	540
Qy	541	LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMskvsVDIFKGFpDEESIVNYTLNQAYQD	600
Db	541	LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMskvsVDIFKGFpDEESIVNYTLNQAYQD	600
Qy	601	NVTVFASVIFQTRKDGSLLPHVHYKIRQNSSFTEKTNEIRRAYWRPGNPtGGRFYFLYGF	660
Db	601	NVTVFASVIFQTRKDGSLLPHVHYKIRQNSSFTEKTNEIRRAYWRPGNPtGGRFYFLYGF	660
Qy	661	VWIQDMMERAIIDTFVGHDVVEPGSYQMFPYPcyTRDDFLFVIEHMmPLCMVISwvysv	720
Db	661	VWIQDMMERAIIDTFVGHDVVEPGSYQMFPYPcyTRDDFLFVIEHMmPLCMVISwvysv	720
Qy	721	AMTIQHIVAeKEHRLKEVMkTMGLNNAvHWAwFITgFVQLsisvtALTAILKYGQVLMH	780
Db	721	AMTIQHIVAeKEHRLKEVMkTMGLNNAvHWAwFITgFVQLsisvtALTAILKYGQVLMH	780
Qy	781	SHVVIIWLFLAVYAVAtIMFCFLSVLYSKAKLASACGGIIYfLSYVPymyVAIREEvAH	840
Db	781	SHVVIIWLFLAVYAVAtIMFCFLSVLYSKAKLASACGGIIYfLSYVPymyVAIREEvAH	840
Qy	841	DKITAFEKCIASLMSTTAfGLGskyfAlYeVAgvgIQWhTfsqSPveGDDFNLLLAVTML	900
Db	841	DKITAFEKCIASLMSTTAfGLGskyfAlYeVAgvgIQWhTfsqSPveGDDFNLLLAVTML	900
Qy	901	MVDAVVYGILTwyIEAvHPgmYGLPRWyfPlQksYwlGsgRteAwEwsWPwARTprLSV	960
Db	901	MVDAVVYGILTwyIEAvHPgmYGLPRWyfPlQksYwlGsgRteAwEwsWPwARTprLSV	960
Qy	961	MEEDQAcAMESRRFEETRGMeEEPTHlPlVVCvdKLTKvyKdDkkLaLNKslNLyENQV	1020
Db	961	MEEDQAcAMESRRFEETRGMeEEPTHlPlVVCvdKLTKvyKdDkkLaLNKslNLyENQV	1020

Qy	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR	1080
Db	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR	1080
Qy	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAFVG	1140
Db	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAFVG	1140
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Db	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Qy	1201	KCCGSPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Db	1201	KCCGSPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Qy	1261	HVASCLLVSDTSTELSYPSEAAKKGAERLRFQHLERSLDALHLSSFGMLDITLLEEVFL	1320
Db	1261	HVASCLLVSDTSTELSYPSEAAKKGAERLRFQHLERSLDALHLSSFGMLDITLLEEVFL	1320
Qy	1321	KVSEEDQSLENSEADVKEERKDVLPAGEPASGEGHAGNLARCSELTQSQASLQSSASSVG	1380
Db	1321	KVSEEDQSLENSEADVKEERKDVLPAGEPASGEGHAGNLARCSELTQSQASLQSSASSVG	1380
Qy	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Db	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Qy	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYHNYTQ	1500
Db	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYHNYTQ	1500
Qy	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Db	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Qy	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS	1620
Db	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS	1620
Qy	1621	GPEMWTSAAPSLPRLVREPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILTITGHNV	1680
Db	1621	GPEMWTSAAPSLPRLVREPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILTITGHNV	1680
Qy	1681	EYLLFTSDRFLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Db	1681	EYLLFTSDRFLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Qy	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFII	1800
Db	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFII	1800
Qy	1801	VAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWMLNLYLPATCCVIIIF	1860
Db	1801	VAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWMLNLYLPATCCVIIIF	1860
Qy	1861	VFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFVFEVPSSAYVFLIVINLFIGITAT	1920

Db	1861	VFDLPAYTSPTNFPVAVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGITAT	1920
Qy	1921	VATFLLQLFEHDKDLKVVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKM	1980
Db	1921	VATFLLQLFEHDKDLKVVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKM	1980
Qy	1981	KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQMPVSTKPVEDDQDVASERQR	2040
Db	1981	KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQMPVSTKPVEDDQDVASERQR	2040
Qy	2041	VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRCLGVRPGECFGLLGVNGAGKTSTFKM	2100
Db	2041	VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRCLGVRPGECFGLLGVNGAGKTSTFKM	2100
Qy	2101	LTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW	2160
Db	2101	LTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW	2160
Qy	2161	KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK	2220
Db	2161	KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK	2220
Qy	2221	ARRFLWNLIIDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG	2280
Db	2221	ARRFLWNLIIDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG	2280
Qy	2281	YMITVRTKSSQSVKDVVRFFNRNFP EAMLERHHTKVQYQLKSEHISLAQVFSKMEQVSG	2340
Db	2281	YMITVRTKSSQSVKDVVRFFNRNFP EAMLERHHTKVQYQLKSEHISLAQVFSKMEQVSG	2340
Qy	2341	VLGIEDYSVSQTTLDNVFNFAKKQSDNLEQQETEPSPALQSPLGCLLSLLRPRSAPTL	2400
Db	2341	VLGIEDYSVSQTTLDNVFNFAKKQSDNLEQQETEPSPALQSPLGCLLSLLRPRSAPTL	2400
Qy	2401	RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC	2436
Db	2401	RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC	2436

RESULT 2

US-09-795-693-8

; Sequence 8, Application US/09795693

; Patent No. US20020068710A1

; GENERAL INFORMATION:

; APPLICANT: Glucksmann, Maria A.

; TITLE OF INVENTION: 20685, 579, 17114, 23821, 33894, and

; TITLE OF INVENTION: 32613, No. US20020068710A1el Human Transporters

; FILE REFERENCE: 35800/209292

; CURRENT APPLICATION NUMBER: US/09/795,693

; CURRENT FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: 60/185,906

; PRIOR FILING DATE: 2000-02-29

; NUMBER OF SEQ ID NOS: 42

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 8

; LENGTH: 2436

; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-795-693-8

Query Match 99.9%; Score 12656; DB 9; Length 2436;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2434; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MGFLHQQLLLWKNVTLKRRSPWVLAFEIFIPLVLFILLGLRQKKPTISVKEVPFYTAA	60
Db	1	MGFLHQQLLLWKNVTLKRRSPWVLAFEIFIPLVLFILLGLRQKKPTISVKEVSFYTAA	60
Qy	61	PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLDRVVEGNLFDPARPSLGSE	120
Db	61	PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLDRVVEGNLFDPARPSLGSE	120
Qy	121	LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSPNSTAQAL	180
Db	121	LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSPNSTAQAL	180
Qy	181	LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLLAPALLEQLTC	240
Db	181	LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLLAPALLEQLTC	240
Qy	241	TPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRESGLSAELRNQLDVAKVSQQL	300
Db	241	TPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRESGLSAELRNQLDVAKVSQQL	300
Qy	301	GLDAPNGSDSSPQAPPPRRLQALLGDLLDAQVQLQDQVDVLSALALLPQGACTGRTPGPP	360
Db	301	GLDAPNGSDSSPQAPPPRRLQALLGDLLDAQVQLQDQVDVLSALALLPQGACTGRTPGPP	360
Qy	361	ASGAGGAANGTGAGAVMGNPATAEEGAPSAALATPDTLQGQCSAFVQLWAGLQPILCGN	420
Db	361	ASGAGGAANGTGAGAVMGNPATAEEGAPSAALATPDTLQGQCSAFVQLWAGLQPILCGN	420
Qy	421	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF	480
Db	421	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF	480
Qy	481	AFVGNVTHYAQVWLNISAEIRSFLEQGRLQQHLRWLQQYVAELRLHPEALNLSLDELPPA	540
Db	481	AFVGNVTHYAQVWLNISAEIRSFLEQGRLQQHLRWLQQYVAELRLHPEALNLSLDELPPA	540
Qy	541	LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPEESIVNYTLNQAYQD	600
Db	541	LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPEESIVNYTLNQAYQD	600
Qy	601	NVTVFASVIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF	660
Db	601	NVTVFASVIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF	660
Qy	661	VWIQDMMERAIIDTFVGHDVVEPGSYVQMFPYPCYTRDDFLFVIEHMMPLCMVISWVYSV	720
Db	661	VWIQDMMERAIIDTFVGHDVVEPGSYVQMFPYPCYTRDDFLFVIEHMMPLCMVISWVYSV	720
Qy	721	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAFITGFVQLSISVTALTALILKYQVLMH	780

Db	721	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAFITGFVQLSISVTALTALIKYQVLIH	780
Qy	781	SHVVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAH	840
Db	781	SHVVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAH	840
Qy	841	DKITAFEKCIASLMSTTAFGLGSKYFALYEAVAGVGIQWHTFSQSPVEGDDFNLLAVTML	900
Db	841	DKITAFEKCIASLMSTTAFGLGSKYFALYEAVAGVGIQWHTFSQSPVEGDDFNLLAVTML	900
Qy	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSV	960
Db	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSV	960
Qy	961	MEEDQACAMESRRFEETRGMEEPTHLPVVCVDKLTKVYKDDKKLALNKLNLNLYENQV	1020
Db	961	MEEDQACAMESRRFEETRGMEEPTHLPVVCVDKLTKVYKDDKKLALNKLNLNLYENQV	1020
Qy	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR	1080
Db	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR	1080
Qy	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAFVG	1140
Db	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAFVG	1140
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRITILLSTHHMDEADLLGDRIAIISHGKL	1200
Db	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRITILLSTHHMDEADLLGDRIAIISHGKL	1200
Qy	1201	KCCGSPLFLKGTYG DYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Db	1201	KCCGSPLFLKGTYG DYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Qy	1261	HVASCLLVSDTSTELSYILPSEAAKKAFAERLFQHLERSLDALHLSSFGLMDDTTLEEVFL	1320
Db	1261	HVASCLLVSDTSTELSYILPSEAAKKAFAERLFQHLERSLDALHLSSFGLMDDTTLEEVFL	1320
Qy	1321	KVSEEDQSLENSEADVKE SRKDVLP GAEGPASGEGHAGNLARCSELTQSQASLQSASSVG	1380
Db	1321	KVSEEDQSLENSEADVKE SRKDVLP GAEGPASGEGHAGNLARCSELTQSQASLQSASSVG	1380
Qy	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Db	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Qy	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQ	1500
Db	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQ	1500
Qy	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Db	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Qy	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDELQAWNVS LPPTA	1620

Db	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDELQAWNVS L PPTA	1620
Qy	1621	GPEMWTSAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILT DITGHNVS	1680
Db	1621	GPEMWTSAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILT DITGHNVS	1680
Qy	1681	EYLLFTSDRFLHRYGAITFGNVLSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Db	1681	EYLLFTSDRFLHRYGAITFGNVLSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Qy	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGT DVVIAIFII	1800
Db	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGT DVVIAIFII	1800
Qy	1801	VAMSFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDM LNYLVPATCCVIILF	1860
Db	1801	VAMSFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDM LNYLVPATCCVIILF	1860
Qy	1861	VFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITAT	1920
Db	1861	VFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITAT	1920
Qy	1921	VATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDM	1980
Db	1921	VATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDM	1980
Qy	1981	KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRM PVSTKPVEDD VDVASERQR	2040
Db	1981	KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRM PVSTKPVEDD VDVASERQR	2040
Qy	2041	VLRGDADNDMVKIENLT KVYKSRKIGRILAVDRLCLGVRPGECFGLLG VNGAGKTSTFKM	2100
Db	2041	VLRGDADNDMVKIENLT KVYKSRKIGRILAVDRLCLGVRPGECFGLLG VNGAGKTSTFKM	2100
Qy	2101	LTGDESTTGGEAFVNGH SVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW	2160
Db	2101	LTGDESTTGGEAFVNGH SVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW	2160
Qy	2161	KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK	2220
Db	2161	KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK	2220
Qy	2221	ARRFLWNLILD LIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG	2280
Db	2221	ARRFLWNLILD LIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG	2280
Qy	2281	YMITVRTKSSQSVKDVVRFFNRNFPEAM LKERHHTKVQYQLKSEHISLAQVFSKMEQVSG	2340
Db	2281	YMITVRTKSSQSVKDVVRFFNRNFPEAM LKERHHTKVQYQLKSEHISLAQVFSKMEQVSG	2340
Qy	2341	VLGIEDYSVSQTTLDNVFNFAKKQSDNLEQQETEP PSALQSPLGCLLSLLRPRSAPTEL	2400
Db	2341	VLGIEDYSVSQTTLDNVFNFAKKQSDNLEQQETEP PSALQSPLGCLLSLLRPRSAPTEL	2400
Qy	2401	RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC	2436
Db	2401	RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC	2436

RESULT 3  
 US-10-156-239-8  
 ; Sequence 8, Application US/10156239  
 ; Publication No. US20030036074A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Glucksmann, Maria A.  
 ; APPLICANT: Kapeller-Libermann, Rosana  
 ; TITLE OF INVENTION: No. US20030036074A1 Nucleic Acid Sequences Encoding  
 Human Transporters, A Human  
 ; TITLE OF INVENTION: ATPase Molecule, A Human Ubiquitin Hydrolase-Like  
 Molecule, A Human  
 ; TITLE OF INVENTION: Ubiquitin Conjugating Enzyme-Like Molecule, and Uses  
 Therefor  
 ; FILE REFERENCE: 35800/247645  
 ; CURRENT APPLICATION NUMBER: US/10/156,239  
 ; CURRENT FILING DATE: 2002-05-24  
 ; PRIOR APPLICATION NUMBER: 09/795,693  
 ; PRIOR FILING DATE: 2001-02-28  
 ; PRIOR APPLICATION NUMBER: 60/185,906  
 ; PRIOR FILING DATE: 2000-02-29  
 ; PRIOR APPLICATION NUMBER: 09/809,557  
 ; PRIOR FILING DATE: 2001-03-15  
 ; PRIOR APPLICATION NUMBER: 60/192,018  
 ; PRIOR FILING DATE: 2000-03-24  
 ; PRIOR APPLICATION NUMBER: 09/808,568  
 ; PRIOR FILING DATE: 2001-03-14  
 ; PRIOR APPLICATION NUMBER: 60/191,790  
 ; PRIOR FILING DATE: 2000-03-24  
 ; PRIOR APPLICATION NUMBER: 09/808,767  
 ; PRIOR FILING DATE: 2001-03-15  
 ; PRIOR APPLICATION NUMBER: 60/191,781  
 ; PRIOR FILING DATE: 2000-03-24  
 ; NUMBER OF SEQ ID NOS: 60  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 8  
 ; LENGTH: 2436  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-156-239-8

Query Match 99.9%; Score 12656; DB 14; Length 2436;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 2434; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MGFLHQQLLLWKNVTLKRRSPWVLAFAEIFIPLVLFILLGLRQKKPTISVKEVPFYTAA	60
Db	1	MGFLHQQLLLWKNVTLKRRSPWVLAFAEIFIPLVLFILLGLRQKKPTISVKEVSFYTAA	60
Qy	61	PLTSAGILPVMQSLCPDQQRDEFGFLQYANSTVTQLLERLDRVVEEGNLFDPARPSLGSE	120
Db	61	PLTSAGILPVMQSLCPDQQRDEFGFLQYANSTVTQLLERLDRVVEEGNLFDPARPSLGSE	120
Qy	121	LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSLPNSTAQAL	180
Db	121	LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSLPNSTAQAL	180

Qy	181	LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPALLEQLTC	240
Db	181	LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPALLEQLTC	240
Qy	241	TPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARFSGLSAELRNQLDVAKVSQQL	300
Db	241	TPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARFSGLSAELRNQLDVAKVSQQL	300
Qy	301	GLDAPNGSDSSPQAPPPRRLQALLGDLLDAQKVLQDQVLSALALLLPQGACTGRTPGPP	360
Db	301	GLDAPNGSDSSPQAPPPRRLQALLGDLLDAQKVLQDQVLSALALLLPQGACTGRTPGPP	360
Qy	361	ASGAGGAANGTGAGAVMGNATAEEGAPSAAALATPDTLQGQCSAFVQLWAGLQPILCGN	420
Db	361	ASGAGGAANGTGAGAVMGNATAEEGAPSAAALATPDTLQGQCSAFVQLWAGLQPILCGN	420
Qy	421	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF	480
Db	421	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF	480
Qy	481	AFVGNVTHYAQVWLNISAEIRSFLEQGRLLQHLRWLQQYVAELRLHPEALNLSLDELPPA	540
Db	481	AFVGNVTHYAQVWLNISAEIRSFLEQGRLLQHLRWLQQYVAELRLHPEALNLSLDELPPA	540
Qy	541	LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQD	600
Db	541	LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQD	600
Qy	601	NVTVFASVIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF	660
Db	601	NVTVFASVIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF	660
Qy	661	VWIQDMMERAIIDTFVGHDVVEPGSYVQMFPYPCYTRDDFLFVIEHMMPLCMVISWVYSV	720
Db	661	VWIQDMMERAIIDTFVGHDVVEPGSYVQMFPYPCYTRDDFLFVIEHMMPLCMVISWVYSV	720
Qy	721	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAFITGFVQLSISVTALTAILKYGQVLMH	780
Db	721	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAFITGFVQLSISVTALTAILKYGQVLIH	780
Qy	781	SHVVIWLFLAVYAVATIMFCFLSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAH	840
Db	781	SHVVIWLFLAVYAVATIMFCFLSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAH	840
Qy	841	DKITAFEKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTML	900
Db	841	DKITAFEKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTML	900
Qy	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSV	960
Db	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSV	960
Qy	961	MEEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTQVYKDDKKLALNKLNLNLYENQV	1020
Db	961	MEEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTQVYKDDKKLALNKLNLNLYENQV	1020



Qy	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRL	1080
Db	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRL	1080
Qy	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVG	1140
Db	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVG	1140
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Db	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Qy	1201	KCCGSPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Db	1201	KCCGSPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Qy	1261	HVASCLLVSDTSTELSYILPSEAAKKGAERLFOHLERSLDALHLSSFGLMDTTLEEVFL	1320
Db	1261	HVASCLLVSDTSTELSYILPSEAAKKGAERLFOHLERSLDALHLSSFGLMDTTLEEVFL	1320
Qy	1321	KVSEEDQSLENSEADVKE SRKDVLPGAEGPASGEGHAGNLARCELTQSQASLQSASSVG	1380
Db	1321	KVSEEDQSLENSEADVKE SRKDVLPGAEGPASGEGHAGNLARCELTQSQASLQSASSVG	1380
Qy	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Db	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Qy	1441	HGLLVKR FHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQ	1500
Db	1441	HGLLVKR FHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQ	1500
Qy	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Db	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Qy	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS LPPTA	1620
Db	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS LPPTA	1620
Qy	1621	GPEMWTSA PSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILT DITGHNVS	1680
Db	1621	GPEMWTSA PSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILT DITGHNVS	1680
Qy	1681	EYLLFTSDRFR L HRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Db	1681	EYLLFTSDRFR L HRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Qy	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHMPMNKTSASLSLDYLLQGT DVVIAIFII	1800
Db	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHMPMNKTSASLSLDYLLQGT DVVIAIFII	1800
Qy	1801	VAMSFVPAS FVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVIILF	1860
Db	1801	VAMSFVPAS FVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVIILF	1860
Qy	1861	VFDLPAYTSPTNFP AVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGITAT	1920

Db	1861	VFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITAT	1920
Qy	1921	VATFLLQLFEHDKDLKVVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDM	1980
Db	1921	VATFLLQLFEHDKDLKVVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDM	1980
Qy	1981	KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVPSTKPVEDDVDVASERQR	2040
Db	1981	KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVPSTKPVEDDVDVASERQR	2040
Qy	2041	VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRCLGVRPGECFGLLGVNGAGKTSTFKM	2100
Db	2041	VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRCLGVRPGECFGLLGVNGAGKTSTFKM	2100
Qy	2101	LTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW	2160
Db	2101	LTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW	2160
Qy	2161	KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK	2220
Db	2161	KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK	2220
Qy	2221	ARRFLWNLILDLIKTRGSVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG	2280
Db	2221	ARRFLWNLILDLIKTRGSVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG	2280
Qy	2281	YMITVRTKSSQSVKDVVRFFNRFPEAMLERHHTKVQYQLKSEHISLAQVFSKMEQVSG	2340
Db	2281	YMITVRTKSSQSVKDVVRFFNRFPEAMLERHHTKVQYQLKSEHISLAQVFSKMEQVSG	2340
Qy	2341	VLGIEDYSVSQTTLDNVFVNFAKKQSDNLEQQETEPSPALQSPLGCLLSLLRPRSAPTEL	2400
Db	2341	VLGIEDYSVSQTTLDNVFVNFAKKQSDNLEQQETEPSPALQSPLGCLLSLLRPRSAPTEL	2400
Qy	2401	RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC	2436
Db	2401	RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC	2436

#### RESULT 4

US-10-199-485-8

; Sequence 8, Application US/10199485

; Publication No. US20030077626A1

; GENERAL INFORMATION:

; APPLICANT: Glucksmann, Maria A.

; APPLICANT: Silos-Santiago, Inmaculada

; TITLE OF INVENTION: 20685, 579, 17114, 23821, 33894, and

; TITLE OF INVENTION: 32613, No. US20030077626A1el Human Transporters

; FILE REFERENCE: 35800/249468

; CURRENT APPLICATION NUMBER: US/10/199,485

; CURRENT FILING DATE: 2002-07-18

; PRIOR APPLICATION NUMBER: 09/795,693

; PRIOR FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: 60/185,906

; PRIOR FILING DATE: 2000-02-29

; NUMBER OF SEQ ID NOS: 42

; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 2436  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-199-485-8

Query Match 99.9%; Score 12656; DB 14; Length 2436;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2434; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 MGFLHQIQLLLWKNVTLKRRSPWVLAFEIFIPLVLFFILLGLRQKKPTISVKEVPFYTAA 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MGFLHQIQLLLWKNVTLKRRSPWVLAFEIFIPLVLFFILLGLRQKKPTISVKEVSFYTAA 60

Qy     61 PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLDRVVEEGNLFDPARPSLGSE 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLDRVVEEGNLFDPARPSLGSE 120

Qy    121 LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSLPNSTAQAL 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSLPNSTAQAL 180

Qy    181 LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLLAPALLEQLTC 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLLAPALLEQLTC 240

Qy    241 TPGSGELGRILTVPEsqKALQGYRDAVCSGQAAARARFSGLSAELRNQLDVAKVSQQL 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TPGSGELGRILTVPEsqKALQGYRDAVCSGQAAARARFSGLSAELRNQLDVAKVSQQL 300

Qy    301 GLDAPNGSDSSPQAPPPRRLQALLGDLLDAQKVLQDQDVLQDQDVLQDQDVLQDQDVL 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GLDAPNGSDSSPQAPPPRRLQALLGDLLDAQKVLQDQDVLQDQDVLQDQDVLQDQDVL 360

Qy    361 ASGAGGAANGTGAGAVMGPNATAEEGAPSAAALATPDTLQGQCSAFVQLWAGLQPILCGN 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ASGAGGAANGTGAGAVMGPNATAEEGAPSAAALATPDTLQGQCSAFVQLWAGLQPILCGN 420

Qy    421 NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF 480

Qy    481 AFVGNVTHYAQVWLNISAEIRSFLQGRLLQHLRWLQYVAELRLHPEALNLSLDELPPA 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AFVGNVTHYAQVWLNISAEIRSFLQGRLLQHLRWLQYVAELRLHPEALNLSLDELPPA 540

Qy    541 LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQD 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQD 600

Qy    601 NVTVFASVIFQTRKDGSPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 NVTVFASVIFQTRKDGSPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF 660

Qy    661 VWIQDMMERAIIDTFVGHQVVEPGSYVQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSV 720
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Db	661	VWIQDMMERAIIDTFVGHDVVEPGSYVQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSV	720
Qy	721	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAWFITGQVQLSISVTALTALILKYQVIMH	780
Db	721	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAWFITGQVQLSISVTALTALILKYQVLIH	780
Qy	781	SHVVIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAH	840
Db	781	SHVVIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAH	840
Qy	841	DKITAFEKCIASLMSTTAFGLGSKYFALYEAVGVGIQWHTFSQSPVEGDDFNLLAVTML	900
Db	841	DKITAFEKCIASLMSTTAFGLGSKYFALYEAVGVGIQWHTFSQSPVEGDDFNLLAVTML	900
Qy	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSV	960
Db	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSV	960
Qy	961	MEEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTQVYKDDKKLALNKLNLNLYENQV	1020
Db	961	MEEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTQVYKDDKKLALNKLNLNLYENQV	1020
Qy	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRL	1080
Db	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRL	1080
Qy	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLGGMKRKLVAIAFVG	1140
Db	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLGGMKRKLVAIAFVG	1140
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Db	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Qy	1201	KCCGSPLFLKGTYG DYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Db	1201	KCCGSPLFLKGTYG DYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Qy	1261	HVASCLLVSDTSTELSYILPSEAAKKGAFLERLQHLERSLDALHLSSFGLMDTTLEEVFL	1320
Db	1261	HVASCLLVSDTSTELSYILPSEAAKKGAFLERLQHLERSLDALHLSSFGLMDTTLEEVFL	1320
Qy	1321	KVSEEDQSLENSEADVKE SRKDVLPAGEGPASGEGHAGNLARCSELTQSQASLQSASSVG	1380
Db	1321	KVSEEDQSLENSEADVKE SRKDVLPAGEGPASGEGHAGNLARCSELTQSQASLQSASSVG	1380
Qy	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Db	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Qy	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVP EIGDLPPLVLSPSQYHNYTQ	1500
Db	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVP EIGDLPPLVLSPSQYHNYTQ	1500
Qy	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Db	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560

Qy	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDELQAWNVS L PPTA	1620
Db	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDELQAWNVS L PPTA	1620
Qy	1621	GPEMWTSA PSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILT DITGHNVS	1680
Db	1621	GPEMWTSA PSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILT DITGHNVS	1680
Qy	1681	EYLLFTSDRFR L HRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Db	1681	EYLLFTSDRFR L HRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Qy	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGT D VVIAIFII	1800
Db	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGT D VVIAIFII	1800
Qy	1801	VAMSFVPAS FV VFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDM L NYLVPATCCVIILF	1860
Db	1801	VAMSFVPAS FV VFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDM L NYLVPATCCVIILF	1860
Qy	1861	VFDLPAYTSPTNFP AVLSLFLLYGWSITPIMYPASFWFEPSSAYVFLIVINLF I G I T A T	1920
Db	1861	VFDLPAYTSPTNFP AVLSLFLLYGWSITPIMYPASFWFEPSSAYVFLIVINLF I G I T A T	1920
Qy	1921	VATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQ FDKM	1980
Db	1921	VATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQ FDKM	1980
Qy	1981	KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQ RMPVSTKPVEDD V D V A S E R Q R	2040
Db	1981	KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQ RMPVSTKPVEDD V D V A S E R Q R	2040
Qy	2041	VLRGDADNDMVKIENLT K VYKSRKIGRILAVDRLCLGVRPGECFGLLG V N G A G K T S T F K M	2100
Db	2041	VLRGDADNDMVKIENLT K VYKSRKIGRILAVDRLCLGVRPGECFGLLG V N G A G K T S T F K M	2100
Qy	2101	LTGDESTTGGEAFVNGH S V L K E L L Q V Q Q S L G Y C P Q C D A L F D E L T A R E H L Q L Y T R L R G I S W	2160
Db	2101	LTGDESTTGGEAFVNGH S V L K E L L Q V Q Q S L G Y C P Q C D A L F D E L T A R E H L Q L Y T R L R G I S W	2160
Qy	2161	KDEARVVKWALEKLELT KYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK	2220
Db	2161	KDEARVVKWALEKLELT KYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK	2220
Qy	2221	ARRFLWN L I L D L I K T G R S V V L T S H S M E E C E A L C T R L A I M V N G R L R C L G S I Q H L K N R F G D G	2280
Db	2221	ARRFLWN L I L D L I K T G R S V V L T S H S M E E C E A L C T R L A I M V N G R L R C L G S I Q H L K N R F G D G	2280
Qy	2281	YMITVRTKSSQSVKDVVRFFN R N F P E A M L K E R H H T K V Q Y Q L K S E H I S L A Q V F S K M E Q V S G	2340
Db	2281	YMITVRTKSSQSVKDVVRFFN R N F P E A M L K E R H H T K V Q Y Q L K S E H I S L A Q V F S K M E Q V S G	2340
Qy	2341	VLGIEDYSVSQT T L D N V F V N F A K K Q S D N L E Q Q E T E P P S A L Q S P L G C L L S L L R P R S A P T E L	2400
Db	2341	VLGIEDYSVSQT T L D N V F V N F A K K Q S D N L E Q Q E T E P P S A L Q S P L G C L L S L L R P R S A P T E L	2400

Qy            2401 RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2436  
               | | | | | | | | | | | | | | | | | | | | | | | |  
Db            2401 RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2436

## RESULT 5

US-10-072-621-8

; Sequence 8, Application US/10072621

; Publication No. US20020169137A1

; GENERAL INFORMATION:

; APPLICANT: Reiner, Peter B.

; APPLICANT: Connop, Bruce P.

; APPLICANT: Pollard, Michelle

; TITLE OF INVENTION: REGULATION OF AMYLOID PRECURSOR PROTEIN EXPRESSION

; TITLE OF INVENTION: BY MODIFICATION OF ABC TRANSPORTER EXPRESSION OR ACTIVITY

: FILE REFERENCE: 100103.402

: CURRENT APPLICATION NUMBER: US/10/072,621

; CURRENT FILING DATE: 2002-02-08

; NUMBER OF SEO ID NOS: 10

```
; SOFTWARE: FastSEO for Windows Version 4.0
```

; SEQ ID NO 8

```
; LENGTH: 2001
```

```
; TYPE: PRT
```

; ORGANISM: Homo sapiens

; FEATURE:

```
; NAME/KEY: VARIANT
```

; LOCATION: 30, 70, 280, 477, 558, 1471, 1651, 1689, 1724

; OTHER INFORMATION: Xaa = Any Amino Acid

; FEATURE:

```
; NAME/KEY: VARIANT
```

; LOCATION: 30, 70, 280, 477, 558, 1471, 1651, 1689, 1724

OTHER INFORMATION: Xaa = Any Amino Acid

US-10-072-621-8

Query Match 80.9%; Score 10249; DB 13; Length 2001;  
Best Local Similarity 98.5%; Pred. No. 0;  
Matches 1973; Conservative 2; Mismatches 26; Indels 2; Gaps 2;

QY 434 MSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVW 493  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 MSSLGFTSKEQRNLGLLVHLMTSNPKILYXPAGSEVDRVILKANETFAFVGNVTHYAQVW 60

QY            494 LNISAEIRSFLEQGRLQQHLRWLQQYVAELRLHPEALNLSLDELPPALRQDNFSLPSGMA       553  
                 |||||     |||||||     |||||||     |||||||     |||||||     |||||||     |||||||     |||||||  
Db            61 LNISAEIRSXLEOGRLOOHRLWLOOYVAELRPHPHALNLSLDELPPALRODNFSLPSGMA       120

Qy 554 LLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQDNVTVFASVIFQTR 613  
 |||  
 Db 121 LLQQLDTIDNAPCGWIOFMSKVSVDIFKGFDPDEESIVNYTLNQAYQDNVTVFAGVIFQTR 180

Qy           614 KDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGFVWIQDMMERAIID 673  
             | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db           181 KDGSLPPHVHYKIRONSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGFVWIODMMERAIID 240

Qy           674 TFGVGHVVPEGSYVQMFPYPCYTRDDFLFVIEHMMPLCMVISWVYSVAMTIQHIVAEKEH   733  
               |||  
               |||

Db 241 TFVGHDVVEPGSYVQMFYPYCYTRDDFLFVIEHMMPLCMXISWVYSVAMTIQHIVAKEH 300

Qy 734 RLKEVMKTMGLNNAVHWVAWFITGFVQLSISVTALTALIKYGQVLMHSHVVIWLFLAVY 793  
 |||

Db 301 RLKEVMKTMGLNNAVHWVAWFITGFVQLSISVTALTALIKYGQVLMHSHVVIWLFLAVY 360

Qy 794 AVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHDKITAFEKCIASL 853  
 |||

Db 361 AVATIMFCFLVSVLYSKAKLASA-GGIIYFLSYVPYMYVAIREEVAHDKITAFEKCIASL 419

Qy 854 MSTTAFGLGSKYFALYEAVAGVGIQWHTFSQSPVEGDDFNLLAVTMLMVDVAVYGILTXY 913  
 |||

Db 420 MSTTAFGLGSKYFALYEAVAGVGIQWHTFSQSPVEGDDFNLLAVTMLMVDVAVYGILXWY 479

Qy 914 IEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSVMEDQACAMESRR 973  
 |||

Db 480 IEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSVMEDQACAMESRR 539

Qy 974 FEETRGMEEEPHTLPLVVCVDKLTKVYKDDKKLALNKLNLNLYENQVVSFLGHNGAGKTT 1033  
 |||

Db 540 FEETRGMEEEPHTLPLVVXVDKLTKVYKDDKKLALNKLNLNLYENQVVSFLGHNGAGKTT 599

Qy 1034 TMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSRLK 1093  
 |||

Db 600 TMSILTGLFPPTSGSATIYGHDIRTEMDEIRKN-GHVPQHNVLFDRLTVEEHLWFYSRLK 658

Qy 1094 SMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVGGSRAIILDEPTAG 1153  
 |||

Db 659 SMAQEEIPREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVGGSRAIILDEPTAG 718

Qy 1154 VDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPLFLKGTY 1213  
 |||

Db 719 VDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPLFLKGTY 778

Qy 1214 GDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHVASCLLVSDTST 1273  
 |||

Db 779 GDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHVASCLLVSDTST 838

Qy 1274 ELSYILPSEAAKKGAFERLFQHLERSLDALHLSSFGLMDTTLEEVLKVSEEDQSLENSE 1333  
 |||

Db 839 ELSYILPSEAAKKGAFERLFQHLERSLDALHLSSFGLMDTTLEEVLKVSGGDQSLNSG 898

Qy 1334 ADVKESRKDVLPAGEGPASGEGHAGNLARCSELTQSQASLQSASSVGSARGDEGAGYTDV 1393  
 |||

Db 899 ADVKESRKDVLPAGEGHASGEGHAGNLARCSELTQSQASLQSASSVGSALGDEGAGYTDV 958

Qy 1394 YGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQFHGLLVKRFHCARR 1453  
 |||

Db 959 YGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQFHGLLVKRFHCARR 1018

Qy 1454 NSKALFSQILLPAFFVCVAMTVALSVEIGDLPPLVLSPSQYHNYTQPRGNFIPYANEER 1513  
 |||

Db 1019 NSKALFSQILLPAFFVCVAMTVALSVEIGDLPPLVLSPSQYHNYTQPRGNFIPYANEER 1078

Qy 1514 REYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFD 1573  
 |||

Db 1079 REYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFD 1138

Qy	1574	SMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDELQAWNVS LPPTAGPEMWTSA PSLPR	1633
Db	1139	SMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDELQAWNVS LPPTAGQEMWTSA PSLPR	1198
Qy	1634	LVREPVRC TCSAQGTGFSC PSSVGGHPPQMRVVTGDILTDITGHN VSEYLLFTSDRFR LH	1693
		:	
Db	1199	LVREPVRC TCSAQGTGFSC PNSVGGHPPQMRVVTGDILTDITGHN VSEYLLFTSDRFR LH	1258
Qy	1694	RYGAI TFGNV LKSIPASFGTRAPP MVRKIAVRRAAQVFYNNKG YHSMPTYLNSLNNAILR	1753
Db	1259	RYGAI TFGNV LKSIPASFGTRAPP MVRKIRCARAAQVFYNNKG YHSMPTYLNSLNNAILR	1318
Qy	1754	ANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGT DVVIAIFIIVAMS FVPASFVVF	1813
Db	1319	ANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGT DVVIAIFIIVAMS FVPASFVVF	1378
Qy	1814	LVAEKSTKAKHLQFVSGCNP I IYWLANYVWDM LNYLVPATCCV IILFVFDLPAYTSPTNF	1873
Db	1379	LVAEKSTKAKHLQFVSGCNP I IYWLANYVWDM LNYLVPATCCV IILFVFDLPAYTSPTNF	1438
Qy	1874	PAVLSLFLLYGWSITPIMYPASFWFEPSSAYVFLIVINLFIGITATVATFLLQLFEHDK	1933
Db	1439	PAVLSLFLLYGWSITPIMYPASFWFEPSSAYXFLIVINLFIGITATVATFLLQLFEHDK	1498
Qy	1934	DLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGL	1993
Db	1499	DLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGL	1558
Qy	1994	VAMAVEGVVGFLLTIMCQYNFLRRPQ RMPVSTKPVEDDDVDVASERQ RVLRGDADNDMVKI	2053
Db	1559	VAMAVEGVVGFLLTIMCQYNFLRRPQ RMPVSTKPVEDDDVDVASERQ RVLRGDADNDMVKI	1618
Qy	2054	ENLTKVYKSRKIGRILAVDR LCLGV RPGE C FGLLVGNAGKTSTFKMLTGDESTTGGEAF	2113
Db	1619	ENLTKVYKSRKIGRILAVDR LCLGV RPGE C FGXLGVNAGKTSTFKMLTGDESTTGGEAF	1678
Qy	2114	VNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEK	2173
Db	1679	VNGHSVLKELXQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGIXWKDEARVVKWALEK	1738
Qy	2174	LELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFI FLDEPTTGMDPKARRFLWNLILDLI	2233
Db	1739	LELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFI FLDEPTTGMDPKARRFLWNLILDLI	1798
Qy	2234	KTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVVRTKSSQSV	2293
Db	1799	KTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVVRTKSSQSV	1858
Qy	2294	KDVVRFFNRNFPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTT	2353
Db	1859	KDVVRFFNRNFPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTT	1918
Qy	2354	LDNVFVNFAKKQSDNLEQQETEPPSALQSP LGCLLSLLRPRSAPTELRALVADEPEDLDT	2413
Db	1919	LDNVFVNFAKKQSDNLEQQETEPPSALQSP LGCLLSLLRPRSAPTELRALVADEPEDLDT	1978



Qy 2414 EDEGLISFEEERAQLSFNTDTLC 2436  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1979 EDEGLISFEEERAQLSFNTDTLC 2001

RESULT 6

US-10-297-022-18

; Sequence 18, Application US/10297022

; Publication No. US20030216310A1

; GENERAL INFORMATION:

; APPLICANT: INCYTE GENOMICS, INC.

; APPLICANT: THORNTON, Michael

; APPLICANT: WALIA, Narinder K.

; APPLICANT: YUE, Henry

; APPLICANT: NGUYEN, Danniell B.

; APPLICANT: LAL, Preeti

; APPLICANT: GANDHI, Ameena R.

; APPLICANT: TRIBOULEY, Catherine M.

; APPLICANT: YAO, Monique G.

; APPLICANT: RAMKUMAR, Jayalaxmi

; APPLICANT: AU-YOUNG, Janice

; APPLICANT: LU, Yan

; APPLICANT: TANG, Y. Tom

; APPLICANT: AZIMZAI, Yalda

; APPLICANT: BRUNS, Christopher M.

; APPLICANT: GRIFFIN, Jennifer A.

; APPLICANT: YANG, Junming

; APPLICANT: BAUGHN, Mariah R.

; APPLICANT: SANJANWALA, Madhu S.

; APPLICANT: RAUMANN, Brigitte E.

; APPLICANT: LEE, Ernestine A.

; APPLICANT: HAFALIA, April

; APPLICANT: GREENE, Barrie D.

; APPLICANT: KHAN, Farrah A.

; APPLICANT: KEARNEY, Liam

; APPLICANT: ELLIOTT, Vicky S.

; APPLICANT: SEILHAMER, Jeffrey J.

; APPLICANT: POLICKY, Jennifer L.

; APPLICANT: BOROWSKY, Mark L.

; APPLICANT: BURFORD, Neil

; APPLICANT: DING, Li

; APPLICANT: LU, Dyung Aina M.

; APPLICANT: HILLMAN, Jennifer L.

; TITLE OF INVENTION: TRANSPORTERS AND ION CHANNELS

; FILE REFERENCE: PI-0109 PCT

; CURRENT APPLICATION NUMBER: US/10/297,022

; CURRENT FILING DATE: 2002-11-25

; PRIOR APPLICATION NUMBER: 60/208,424; 60/209,001; 60/210,588; 60/212,335;  
60/213,747; 60/215,391

; PRIOR FILING DATE: 2000-05-26; 2000-06-01; 2000-06-08; 2000-06-16; 2000-06-  
22; 2000-06-29

; NUMBER OF SEQ ID NOS: 54

; SOFTWARE: PERL Program

; SEQ ID NO 18

; LENGTH: 1771

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030216310A1 2311751CD1  
US-10-297-022-18

Query Match / 72.9%; Score 9237; DB 15; Length 1771;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1771; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	666	MMERAIIDTFVGHDVVEPGSYVQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSVAMTIQ	725
Db	1	MMERAIIDTFVGHDVVEPGSYVQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSVAMTIQ	60
Qy	726	HIVAEKEHRLKEVMKTMGLNNVHVVAVFITGFVQLSISVTALTAILKYGQVLMHSHVVI	785
Db	61	HIVAEKEHRLKEVMKTMGLNNVHVVAVFITGFVQLSISVTALTAILKYGQVLMHSHVVI	120
Qy	786	IWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIYFLSYVPYMYVAIREEVAHDKITA	845
Db	121	IWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIYFLSYVPYMYVAIREEVAHDKITA	180
Qy	846	FEKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTMLMVDVAV	905
Db	181	FEKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTMLMVDVAV	240
Qy	906	VYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSVMEEDQ	965
Db	241	VYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSVMEEDQ	300
Qy	966	ACAMESRRFEETRGMEEEPHTLPLVVCVDKLTKVYKDDKKLALNKLSLNLYENQVVSFLG	1025
Db	301	ACAMESRRFEETRGMEEEPHTLPLVVCVDKLTKVYKDDKKLALNKLSLNLYENQVVSFLG	360
Qy	1026	HNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEH	1085
Db	361	HNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEH	420
Qy	1086	LWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVGGSRAI	1145
Db	421	LWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVGGSRAI	480
Qy	1146	ILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGS	1205
Db	481	ILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGS	540
Qy	1206	PLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHVASC	1265
Db	541	PLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHVASC	600
Qy	1266	LLVSDTSTELSYILPSEAAKKGAFAERLFQHLERSLDALHLSSFGLMDTTLEEVFLKVSEE	1325
Db	601	LLVSDTSTELSYILPSEAAKKGAFAERLFQHLERSLDALHLSSFGLMDTTLEEVFLKVSEE	660
Qy	1326	DQSLENSEADVKESSRKDVLPAGEGPASGEGHAGNLARCSELTQSQASLQSASSVGSARGD	1385
Db	661	DQSLENSEADVKESSRKDVLPAGEGPASGEGHAGNLARCSELTQSQASLQSASSVGSARGD	720

Qy	1386	EGAGYTDVYGDIRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQFHGLLV	1445
Db	721	EGAGYTDVYGDIRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQFHGLLV	780
Qy	1446	KRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQPRGNF	1505
Db	781	KRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQPRGNF	840
Qy	1506	IPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSGESR	1565
Db	841	IPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSGESR	900
Qy	1566	LLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPEDDLQAWNVS LPPTAGPEMW	1625
Db	901	LLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPEDDLQAWNVS LPPTAGPEMW	960
Qy	1626	TSAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTDTGHNVSEYLLF	1685
Db	961	TSAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTDTGHNVSEYLLF	1020
Qy	1686	TSDRFLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYSMPTYLN	1745
Db	1021	TSDRFLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYSMPTYLN	1080
Qy	1746	SLNNAILRANLPKSKGNPAAYGITVTNHMPMNKTSASLSLDYLLQGTDVVIAIFIIVAMSF	1805
Db	1081	SLNNAILRANLPKSKGNPAAYGITVTNHMPMNKTSASLSLDYLLQGTDVVIAIFIIVAMSF	1140
Qy	1806	VPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNVLVPATCCVILFVFDLP	1865
Db	1141	VPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNVLVPATCCVILFVFDLP	1200
Qy	1866	AYTSPTNFPVAVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITATVATFL	1925
Db	1201	AYTSPTNFPVAVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITATVATFL	1260
Qy	1926	LQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFE	1985
Db	1261	LQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFE	1320
Qy	1986	WDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDQDVASERQVRVLRGD	2045
Db	1321	WDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDQDVASERQVRVLRGD	1380
Qy	2046	ADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLGVNAGKTSTFKMLTGDE	2105
Db	1381	ADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLGVNAGKTSTFKMLTGDE	1440
Qy	2106	STTGGEAFVNGHSLVKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEAR	2165
Db	1441	STTGGEAFVNGHSLVKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEAR	1500
Qy	2166	VVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFL	2225
Db	1501	VVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFL	1560
Qy	2226	WNLILDLIKTRGSVVLTSMSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITV	2285

```

Db      1561 WNLILDLIKTRGSVVLTSHEMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITV 1620
Qy      2286 RTKSSQSVKDVVRFFNRFPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIE 2345
Db      1621 RTKSSQSVKDVVRFFNRFPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIE 1680
Qy      2346 DYSVSQTTLDNVFNFAKKQSDNLEQQETEPPSALQSPLGCLLSLLRPRSAPTELRALVA 2405
Db      1681 DYSVSQTTLDNVFNFAKKQSDNLEQQETEPPSALQSPLGCLLSLLRPRSAPTELRALVA 1740
Qy      2406 DEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2436
Db      1741 DEPEDLDTEDEGLISFEEERAQLSFNTDTLC 1771

```

RESULT 7

US-10-340-097-119

; Sequence 119, Application US/10340097

; Publication No. US20030162276A1

; GENERAL INFORMATION:

; APPLICANT: Rattner, Amir

; APPLICANT: Sun, Hui

; APPLICANT: Lupski, James R.

; APPLICANT: Nathans, Jeremy

; APPLICANT: Anderson, Kent L.

; APPLICANT: Leppert, Mark

; APPLICANT: Dean, Michael

; APPLICANT: Singh, Nanda

; APPLICANT: Shroyer, No. US20030162276A1h F.

; APPLICANT: Smallwood, Philip M.

; APPLICANT: Allikmets, Rando

; APPLICANT: Lewis, Richard A.

; APPLICANT: Li, Yixin

; TITLE OF INVENTION: Nucleic Acid And Amino Acid Sequences For ATP-Binding Cassette

; TITLE OF INVENTION: Transporter And Methods Of Screening For Agents That Modify ATP-Binding Cassette

; TITLE OF INVENTION: Transporter

; FILE REFERENCE: BYLR0065

; CURRENT APPLICATION NUMBER: US/10/340,097

; CURRENT FILING DATE: 2003-01-10

; PRIOR APPLICATION NUMBER: US/09/032,438A

; PRIOR FILING DATE: 1998-02-27

; PRIOR APPLICATION NUMBER: 60/039,388

; PRIOR FILING DATE: 1997-02-27

; NUMBER OF SEQ ID NOS: 120

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 119

; LENGTH: 1472

; TYPE: PRT

; ORGANISM: Mouse

US-10-340-097-119

Query Match 56.2%; Score 7117; DB 14; Length 1472;

Best Local Similarity 94.2%; Pred. No. 0;

Matches 1388; Conservative 22; Mismatches 60; Indels 4; Gaps 4;

Qy	965	QACAMESRRFEETRGMEEETHLPLVVCVDKLTQVYKDDKKLALNKLNLNLYENQVVSFL	1024
Db	1	QACAMESRRFEETRGMEEETHLPLVVCVDKLTQVYKDDKKLALNKLNLNLYENQVVSFL	60
Qy	1025	GHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEE	1084
Db	61	GHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEE	120
Qy	1085	HLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVGGSSRA	1144
Db	121	HLWFYSRLKSMAQEEIRKETDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVGGSSRA	180
Qy	1145	IILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCG	1204
Db	181	IILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCG	240
Qy	1205	SPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHAVAS	1264
Db	241	SPLFLKGAYKDGYYRLTLVKQPAEPGTSQEPGLASSPGCPRLSSCSEPQVSQFIRKHAVAS	300
Qy	1265	CLLVSDTSTELSILPSEAAKKGAFLERLQHLERSLDALHLSSFGLMDTTLEEVFLKVSE	1324
Db	301	SLLVSDTSTELSILPSEAVKKGAFLERLQQLERSLDALHLSSFGLMDTTLEEVFLKVSE	360
Qy	1325	EDQSLENSEADVKESSRKDVLPAGEPASGEGHAGNLARCSELTQSQASLQSASSVGSARG	1384
Db	361	EDQSLENSEADVKESSRKDVLPAGEGLTAVGGQAGNLARCSELAQSQASLQSASSVGSARG	420
Qy	1385	DEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALS RVQGSRKLDGGWLKVRQFHGLL	1444
Db	421	EEGTGYSDDYGDYRPLFDNLQDPDNVSLQEAEMEALAQVQGSRKLEGWLLKMRQFHGLL	480
Qy	1445	VKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYHNYTQPRGN	1504
Db	481	VKRFHCARRNSKALCSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYHNYTQPRGN	540
Qy	1505	FIPYANEERREYRLRLSPDASPPQQLVSTFRLPSGVGATCVLKSPANGSLGPTNLSSGES	1564
Db	541	FIPYANEERQEYRLRLSPDASPPQQLVSTFRLPSGVGATCVLKSPANGSLGPTNLSSGES	600
Qy	1565	RLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDED-LQAWNVS LPPTAGPE	1623
Db	601	RLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPVKPDEDSLQAWNVS LPPTAGPE	660
Qy	1624	MWTSAPSLPRLVREPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILTDTGHNVS EYL	1683
Db	661	TWTSAPSLPRLVHEPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILTDTGHNVS EYL	720
Qy	1684	LFTSDRFRHLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSMPTY	1743
Db	721	LFTSDRFRHLHRYGAITFGNVQKSIPASFGARVPPMVRKIAVRRVAQVLYNNKGYHSMPTY	780
Qy	1744	LNSLNNAILRANLPKSKGNPAAYGKITVTNHPMNKTSASLSLDYLLQGTDVVIAIFIIVAM	1803
Db	781	LNSLNNAILRANLPKSKGNPAAYKITVTNHPMNKTSASLSLDYLLQGTDVVIAIFIIVAM	840

Qy	1804	SFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMLNYLVPATCCVIILFVFD	1863
Db	841	SFVPASFVVFLVAEKSTKAKHLQFVSGCNPIYWLANYVWDMLNYLVPATCCVIILFVFD	900
Qy	1864	LPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGITATVAT	1923
Db	901	LPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGITATVAT	960
Qy	1924	FLLQLFEHDKDLKVVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSP	1983
Db	961	FLLQLFEHDKDLKVVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSP	1020
Qy	1984	FEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMFVSTKPVEDDDVDASERQRVLR	2043
Db	1021	FEWDIVTRGLVAMTVEGFVGFLLTIMCQYNFLRQPQRLPVSTKPVEDDDVDASERQRVLR	1080
Qy	2044	GDADNDMVKIENLTQVYKSRKIGRILAVDRLCLGV-RPGECFGLLVNGAGKTSTFKMLT	2102
Db	1081	GDADNDMVKIENLTQVYKSRKIGRILAVDRLCLGVCVPGECFGLLVNGAGKTSTFKMLT	1140
Qy	2103	GDESTTGGEAFVNGHSVLEKLLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKD	2162
Db	1141	GDESTTGGEAFVNGHSVLEKLLQVQQSLGYCPQFDPVDELTAAREHLQLYTRLRCIPWKD	1200
Qy	2163	EARVVKWALEKLELTQYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKAR	2222
Db	1201	EAQVVKWALEKLELTQYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKAR	1260
Qy	2223	RFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYM	2282
Db	1261	RFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLHCLGSIQHLKNRFGDGYM	1320
Qy	2283	ITVRTKSSQSVKDVVRFFNRFPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVL	2342
Db	1321	ITVRTKSSQNVKDVVRFFNRFPEAHAQGKTPYKVQYQLKSEHISLAQVFSKMEQVVGVL	1380
Qy	2343	GIEDYSVSQTTLDNVFVNFAKKQSDNLEQQETEPSSALQSPLGCLLSLLRPRSAPTLRA	2402
Db	1381	GIEDYSVSQTTLDNVFVNFAKKQSDNVEQQEAE-PSLPSPLG-LLSLLRPRSAPTLRA	1438
Qy	2403	LVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC	2436
Db	1439	LVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC	1472

RESULT 8

US-10-336-215-119

; Sequence 119, Application US/10336215

; Publication No. US20030170852A1

; GENERAL INFORMATION:

; APPLICANT: Allikments, Rando

; APPLICANT: Anderson, Kent L.

; APPLICANT: Dean, Michael

; APPLICANT: Leppert, Mark

; APPLICANT: Lewis, Richard A.

; APPLICANT: Li, Yixin

; APPLICANT: Lupski, James R.



Db	421	EEGTGYSDGYGDIRPLFDNLQDPDNVSLQEAEMEALAQVGQGSRKLEGWWLKMRFHGLL	480
Qy	1445	VKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQPRGN	1504
Db	481	VKRFHCARRNSKALCSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQPRGN	540
Qy	1505	FIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGES	1564
Db	541	FIPYANEERQEYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGES	600
Qy	1565	RLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDE-LQAWNVSIPPTAGPE	1623
Db	601	RLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPVKPDEDSLQAWNMSIPPTAGPE	660
Qy	1624	MWTSAPSLPRLVREPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILTIDITGHNVSLEYL	1683
Db	661	TWTSAPSLPRLVHEPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILTIDITGHNVSLEYL	720
Qy	1684	LFTSDRFRHLHRYGAITFGNVLSKIPASFGTRAPPMVRKIAVRRRAAQVFYNNKGYHSMPTY	1743
Db	721	LFTSDRFRHLHRYGAITFGNVQKSI PASFGARVPPMVRKIAVRRVAQVLYNNKGYHSMPTY	780
Qy	1744	LNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFIIVAM	1803
Db	781	LNSLNNAILRANLPKSKGNPAAYKITVTNHPMNKTSASLSLDYLLQGTDVVIAIFIIVAM	840
Qy	1804	SFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNVLVPATCCVIIILFVFD	1863
Db	841	SFVPASFVVFLVAEKSTKAKHLQFVSGCNPIYWLANYVWDMNLNVLVPATCCVIIILFVFD	900
Qy	1864	LPAYTSPTNFPVAVLSLFLLYGWSITPIMYPASFWEVFPSSAYVFLIVINLFIGITATVAT	1923
Db	901	LPAYTSPTNFPVAVLSLFLLYGWSITPIMYPASFWEVFPSSAYVFLIVINLFIGITATVAT	960
Qy	1924	FLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDMKSP	1983
Db	961	FLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDMKSP	1020
Qy	1984	FEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDDVDVASERQRVLR	2043
Db	1021	FEWDIVTRGLVAMTVEGFVGFLLTIMCQYNFLRQPQRLPVSTKPVEDDDVDVASERQRVLR	1080
Qy	2044	GDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGV-RPGECEGGLLVNGAGKTSTFKMLT	2102
Db	1081	GDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVCVPGECEGGLLVNGAGKTSTFKMLT	1140
Qy	2103	GDESTTGGEAFVNGHSLVKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKD	2162
Db	1141	GDESTTGGEAFVNGHSLVKDLLQVQQSLGYCPQFDVPDELTAREHLQLYTRLRCIPWKD	1200
Qy	2163	EARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKAR	2222
Db	1201	EAQVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKAR	1260
Qy	2223	RFLWNLILDLIKTRGSVVLTSHEEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYM	2282



Db 1261 RFLWNLILDLIKTRGSVVLTSMSMEECEALCTRLAIMVNGRLHCLGSIQHLKNRFGDGYM 1320

Qy 2283 ITVRTKSSQSVKDVVRFFNRFPEAMLERHHTKVQYQLKSEHISLAQVFSKMEQVSGVL 2342  
 |||||:||||| : : ||||| |||

Db 1321 ITVRTKSSQNVKDVVRFFNRFPEAHAQGKTPYKVQYQLKSEHISLAQVFSKMEQVVGVL 1380

Qy 2343 GIEDYSVSQTTLDNVFVNFAKKQSDNLEQQETEPPSALQSPLGCLLSLLRPRSAPTELRA 2402  
 |||||:|||| | ||:| ||| ||||| |||||

Db 1381 GIEDYSVSQTTLDNVFVNFAKKQSDNVEQQEAE-PSSLPSPLG-LLSLLRPRPAPTELRA 1438

Qy 2403 LVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2436  
 |||||

Db 1439 LVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 1472

RESULT 9

US-10-336-219-119

; Sequence 119, Application US/10336219

; Publication No. US20030170853A1

; GENERAL INFORMATION:

; APPLICANT: Allikmets, Rando

; APPLICANT: Anderson, Kent L.

; APPLICANT: Dean, Michael,

; APPLICANT: Leppert, Mark

; APPLICANT: Lewis, Richard A.

; APPLICANT: Li, Yixin

; APPLICANT: Lupski, James R.

; APPLICANT: Nathans, Jeremy

; APPLICANT: Rattner, Amir

; APPLICANT: Shroyer, No. US20030170853A1h F.

; APPLICANT: Singh, Nanda

; APPLICANT: Smallwood, Philip

; APPLICANT: Sun, Hui

; TITLE OF INVENTION: Methods Of Gene Therapy Using Nucleic Acid Sequences For

; TITLE OF INVENTION: ATP-Binding Cassette Transporter

; FILE REFERENCE: BYLR0072

; CURRENT APPLICATION NUMBER: US/10/336,219

; CURRENT FILING DATE: 2003-01-03

; PRIOR APPLICATION NUMBER: 60/039,388

; PRIOR FILING DATE: 1997-02-27

; PRIOR APPLICATION NUMBER: 09/032,438

; PRIOR FILING DATE: 1998-02-27

; NUMBER OF SEQ ID NOS: 120

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 119

; LENGTH: 1472

; TYPE: PRT

; ORGANISM: Mouse

US-10-336-219-119

Query Match 56.2%; Score 7117; DB 14; Length 1472;

Best Local Similarity 94.2%; Pred. No. 0;

Matches 1388; Conservative 22; Mismatches 60; Indels 4; Gaps 4;

Qy 965 QACAMESRRFEETRGMEEPTHLPVVCVDKLTQVYKDDKKLALNKLNLNLYENQVVSFL 1024  
 ||||| : |||||

Db 1 QACAMESRHFEETRGMEEPTHLPVVCVDKLTQVYKNDKKLALNKLNLNLYENQVVSFL 60

Qy	1025	GHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEE	1084
Db	61	GHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEE	120
Qy	1085	HLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVGGSSRA	1144
		:	
Db	121	HLWFYSRLKSMAQEEIRKETDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVGGSSRA	180
Qy	1145	IILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCG	1204
Db	181	IILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCG	240
Qy	1205	SPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHAVAS	1264
		:	
Db	241	SPLFLKGAYKDGYRLTLVKQPAEPGTSQEPGLASSPSGCPRLSSCSEPQVSQFIRKHAVAS	300
Qy	1265	CLLVSDTSTELSYILPSEAAKKGAFAERLFQHLERSLDALHLSSFGLMDTTLEEVLKVSE	1324
Db	301	SLLVSDTSTELSYILPSEAVKKGAFAERLFQQLHSLDALHLSSFGLMDTTLEEVLKVSE	360
Qy	1325	EDQSLENSEADVKE SRKDVLPGAEGPASGEGHAGNLARCSELTQSQASLQSASSVGSARG	1384
		:	
Db	361	EDQSLENSEADVKE SRKDVLPGAEGLTAVGGQAGNLARCSELAQSQASLQSASSVGSARG	420
Qy	1385	DEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQFHGLL	1444
		:     :	
Db	421	EEGTGYSDGYGDYRPLFDNLQDPDNVSLQEAEAEALAQVGQGSRKLEGWWLKMRQFHGLL	480
Qy	1445	VKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQPRGN	1504
Db	481	VKRFHCARRNSKALCSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQPRGN	540
Qy	1505	FIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTNLNLSSGES	1564
		:	
Db	541	FIPYANEERQEYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTNLNLSSGES	600
Qy	1565	RLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDED-LQAWNVS LPPTAGPE	1623
Db	601	RLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPVKPDEDSLQAWNMS LPPTAGPE	660
Qy	1624	MWTSAPSLPRLVREPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILTDITGHNVSEYL	1683
Db	661	TWTSAPSLPRLVHEPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILTDITGHNVSEYL	720
Qy	1684	LFTSDRFRHLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVEYNNKGYHSMPY	1743
Db	721	LFTSDRFRHLHRYGAITFGNVQKSIPASFGARVPPMVRKIAVRRVAQVLYNNKGYHSMPY	780
Qy	1744	LNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFIIVAM	1803
Db	781	LNSLNNAILRANLPKSKGNPAAYKITVTNHPMNKTSASLSLDYLLQGTDVVIAIFIIVAM	840
Qy	1804	SFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVIIIFVFD	1863
		:	
Db	841	SFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVIIIFVFD	900

Qy	1864	LPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGITATVAT	1923
Db	901	LPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGITATVAT	960
Qy	1924	FLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSP	1983
Db	961	FLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSP	1020
Qy	1984	FEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMFVSTKPVEDDDVDASERQRVLR	2043
Db	1021	FEWDIVTRGLVAMTVEGFVGFLLTIMCQYNFLRQPQLPVSTKPVEDDDVDASERQRVLR	1080
Qy	2044	GDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGV-RPGECEGLLGVNGAGKTSTFKMLT	2102
Db	1081	GDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVCVPGECEGLLGVNGAGKTSTFKMLT	1140
Qy	2103	GDESTTGGEAFVNGHSLVKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKD	2162
Db	1141	GDESTTGGEAFVNGHSLVKDLLQVQQSLGYCPQFDVPVDELTAAREHLQLYTRLRCIPWKD	1200
Qy	2163	EARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKAR	2222
		:	
Db	1201	EAQVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKAR	1260
Qy	2223	RFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYM	2282
Db	1261	RFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLHCLGSIQHLKNRFGDGYM	1320
Qy	2283	ITVRTKSSQSVKDVVRFFNRNFPPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVL	2342
		:      : :	
Db	1321	ITVRTKSSQNVKDVVRFFNRNFPPEAHAGKTPYKVQYQLKSEHISLAQVFSKMEQVVGVL	1380
Qy	2343	GIEDYSVSQTTLDNVFVNFPAKQSDNLEQQETEPSPALQSPLGCLLSLLRPRAPTELRA	2402
		:       :	
Db	1381	GIEDYSVSQTTLDNVFVNFPAKQSDNVEQQEAE-PSSLPSPLG-LLSLLRPRAPTELRA	1438
Qy	2403	LVADPEPDLDETEDEGLISFEEERAQLSFNTDTLC	2436
Db	1439	LVADPEPDLDETEDEGLISFEEERAQLSFNTDTLC	1472

RESULT 10

US-10-617-334-1

; Sequence 1, Application US/10617334

; Publication No. US20040058869A1

; GENERAL INFORMATION:

; APPLICANT: Hayden, Michael R.

; APPLICANT: Brooks-Wilson, Angela R.

; TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING CHOLESTEROL LEVELS

; FILE REFERENCE: 760050-91

; CURRENT APPLICATION NUMBER: US/10/617,334

; CURRENT FILING DATE: 2003-07-10

; PRIOR APPLICATION NUMBER: US 09/526,193

; PRIOR FILING DATE: 2000-03-15

; PRIOR APPLICATION NUMBER: 60/124,702

; PRIOR FILING DATE: 1999-03-15

; PRIOR APPLICATION NUMBER: 60/138,048

; PRIOR FILING DATE: 1999-06-08  
 ; PRIOR APPLICATION NUMBER: 60/139,600  
 ; PRIOR FILING DATE: 1999-06-17  
 ; PRIOR APPLICATION NUMBER: 60/151,977  
 ; PRIOR FILING DATE: 1999-09-01  
 ; NUMBER OF SEQ ID NOS: 287  
 ; SOFTWARE: PatentIn 3.0  
 ; SEQ ID NO 1  
 ; LENGTH: 2261  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-617-334-1

Query Match 33.5%; Score 4240.5; DB 12; Length 2261;  
 Best Local Similarity 39.8%; Pred. No. 0;  
 Matches 1000; Conservative 345; Mismatches 730; Indels 435; Gaps 61;

Qy	6	QLQLLLWKNVTLKRRSPWVLAFEIFIPLVLFFILLGLRQKKPTISVKEVPFYTAAPL TSA	65
		:           :   :       :       :       :	
Db	6	QLRLLLWKNLTFRRRQTCQLLLEVAWPLFIFLILISVRLSYPPYEQHECHF NKA-MPSA	64
Qy	66	GILPVMQSLCPDGQRDEFGFL-----QYANSTVTQLLERLDRVVEEGNLFDPARP	115
		:   : :   : :     :   :   :   :	
Db	65	GTLPVWQGIICNANNPCFRYPTPGEAPGVVGNFNKSIVARLFS DARRLL----LYSQKDT	120
Qy	116	SLGSELEALR--QHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNL SLP	173
		: :       :   : :   :	
Db	121	SMKDMRKVLR TLQQIKKSSSNLKLQDFLVDNETFSG-----FLYHNLSLP	165
Qy	174	NSTAQALLAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELL LPA	233
		:       :   :         :     :     :   :	
Db	166	KSTVDKMLRADV----ILHKVFLQGYQLHLTS-LCNGS-----KSEEMI----	204
Qy	234	LLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAELRNQ L DV	293
		: : :   :   :   :     :   :	
Db	205	---QL---GDQEVSEL CGLPREKLAAAE-----RVLRSNMDI	235
Qy	294	AK-VSQQGLGDAPNGSDSSPQAPPPRRLQALLGDLLD-----AQKVLQD VDVLS	341
		: :   :     :   :         :   : :	
Db	236	LKPILRTL NSTSPFP SKELAEA--TKTLLHSLGTLAQELFSMRSWSDMRQEV MFLT NVNS	293
Qy	342	ALALLLPQGACTGRTPGPPASGAGGAAN-----GTGAGAVMGP NATAEEGAPSAAALATP	396
		: :   :       : :   :       :   :	
Db	294	SSSSTQIYQAVSRIVCGHPEGGLKIKSLNWYEDNNYKALFGGNGTEEDAETFYDN STTP	353
Qy	397	---DTLQGQCSAFVQ--LWAGLQPILCGNNRTIEPEALRRGNMSSLGFTSKEQRNLGL LV	451
		: :   : : :   :   :	
Db	354	YCNDLMKNLESSPLSRIIWKALKPLL VG-----	381
Qy	452	HLMTSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVWLNISAEIRS FLEQGRLQQ	511
		:   : :   :   : :   :   :   :	
Db	382	-----KILYTPDTPATRQVMAEVNKT FQELAVFHDLEGMWEELSPKIWTFMENSQEMD	434
Qy	512	HLRWL-----QQYVAELRLHPE---ALNLSLDELPPALRQDNFS	547
		:                   :   :   :	
Db	435	LVRMLLDSRDNDHFWEQQLDGLDWT AQDIVAFLAKHPEDVQSSNGSVYTWREAFNETN--	492

Qy	548	LPSGMALLQQLDITDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQDNVTVFAS	607
Db	493	-----QAIRTIS-----RFMECVNLNKLEPIATEVWLINKSME--LLDERKFWAG	535
Qy	608	VIFQTRKDGS--LPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTG---GRFYFLYGFVW	662
Db	536	IVFTGITPGSIELPHHVYKIRMDIDNVERTNKIKDGYWDPGRADPFEDMRVWGGGFAY	595
Qy	663	IQDMMERAIIDTFVGHDDVVEPGSYVQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSVAM	722
Db	596	LQDVVEQAIIRVLGTGE-KKTGVYMQQMPYPCYVDDIFLRVMSRSMPLFMTLAWIYSVAV	654
Qy	723	TIQHIVAEKEHRLKEVMKTMGLNNVHVWVAVFITGFVQLSISVTALTAILKYQVLMHSH	782
Db	655	IIKGIVYEKEARLKETMRIMGLDNSILWFSWFISSLIPLVLSAGLLVVLKLGNNLPYSD	714
Qy	783	VWIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHDK	842
Db	715	PSVVFVFLSVFAVVTILQCFLISTLFSRANLAAACGGIIYFTLYLPYVLC-----VAWQD	769
Qy	843	ITAFE-KCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTMLM	901
Db	770	YVGFTLKFIFASLLSPVAFGFGCEYFALFEEQGIGVQWQDNLFESPVEEDGFNLTTSVSMML	829
Qy	902	VDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSVM	961
Db	830	FDTFLYGVMTWYIEAVFPQYGI PRPWYF PCTKSYWFG-----ESDEKSHPGSNQKRIS--	884
Qy	962	EEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTKVYKDDKKLALNKLNLNLYENQVV	1021
Db	885	---EIC-----MEEEPHTLKLGVSIQNLVKVYRDGMKVAVDGLALNFYEGQIT	929
Qy	1022	SFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLT	1081
Db	930	SFLGHNGAGKTTTMSILTGLFPPTSGTAYILGKDIRSEMSTIRQNLGVCQPQHNVLFDMLT	989
Qy	1082	VEEHLWFYSRLKSMAQEEIRREMDKMIEDLEL-SNKRHSLVQTLSSGGMKRKLSVAIAFVG	1140
Db	990	VEEHIWFYARLKLSEKHVKAEMEQLMALDVGLPSSSKLKSQTSQSSGMQRKLSVALAFVG	1049
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHMHDEADLLGDRIAIISHGKL	1200
Db	1050	GSKVVILDEPTAGVDPYSRRGIWELLKYRQGRITILLSTHMHDEADVLGDRIAIISHGKL	1109
Qy	1201	KCCGSPLFLKGTYGDRYRLTLVKRPAEPG-----GPQEPGLAS	1238
Db	1110	CCVGSFLKQNLGTGYLTLVKKDVESSLSSCRNSSSTVSYLKKEDSVSQSSSDAGLGS	1169
Qy	1239	SPPGRAPLSSCSELQVSQFIRKHVASCLLVSDTSTELSYILPSEAAKGAERLRFQHLER	1298
Db	1170	DHESDTLTIDVS--AISNLIRKHVSEARLVEDIGHELTYPYEAKEGAFVELFHEIDD	1227
Qy	1299	SLDALHLSSFLGMDTTLEEVFLKVSEEDQSLENSEADVKESSRKDVLPGAEGPASGEGHAG	1358
Db	1228	RLSDLGISSYGISETTLEEIFLKVAEE-----SGVDA-ETSDGTLP-----	1267
Qy	1359	NLARCSELTQSQASLQSSVGSARGDEGAGYTDVYGDYRPLF-DNPQDPD--NVSLOEV	1415

Db	1268	-----ARRNRRRA-FGDKQSCLRPFTEDDAADPNDSDIDPESR	1303
Qy	1416	EAEALSRV-GQGSRKLDGGWLKVRQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMT	1474
Db	1304	ETDLLSGMDGKGSYQVKGWKLTQQQFVALLWKRLLIARRSRKGFFAQIVLPAVFVCIALV	1363
Qy	1475	VALSVPEIGDLPLVLSPSQYH-NYTQPRGNFIPYANEERREYRLRLSPDASPOQLVSTF	1533
Db	1364	FSLIVPPFGKYPSLELQPWMYNEQYT-----FVSNDAP-----DTGTLELLNAL	1408
Qy	1534	RLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPP	1593
Db	1409	TKDPGFGTRCM-----EGNPI-----	1424
Qy	1594	PSPAPSDSPASPEDLQAWNVSLEPPTAGPEMWTSAPSLPRLVREPVR-----C	1641
Db	1425	-----PD-----TPCQAGEEWTAP-VPQTIMDLFQNGNWTMNPSPAC	1463
Qy	1642	TCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILTDITGHNVSLEYLLFTSDRF-----	1690
Db	1464	QCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLKTYVQIIAKSLKN	1523
Qy	1691	----RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK-----	1721
Db	1524	KIWNVEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTG	1583
Qy	1722	IARRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSAS	1781
Db	1584	LDTRNNVKVWFNNKGWHAISSFLNVINNAAILRANLQKGE-NPSHYGITAFNHPLNLTKQQ	1642
Qy	1782	LS-LDYLLQGTDVVIAIFIIVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLAN	1840
Db	1643	LSEVALMTTSVDVLVSICVIFAMSFVPASFVFLIQERVSKAKHLQFISGVKPVYIWLNS	1702
Qy	1841	YVWDMNLNVLVPATCCVILFVFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFVFEV	1900
Db	1703	FVWDMCNVVPATLVIIIFICFQQKSYVSSTNLPVLALLLLLYGWSITPLMYPASFVEKI	1762
Qy	1901	PSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGIME	1960
Db	1763	PSTAYVVLTSVNLFIGINGSVATFVLELEFT-DNKLNNINDILKSVFLIFPHFCLGRGLID	1821
Qy	1961	MAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQR	2020
Db	1822	MVKNQAMADALERFGE-NRFVSPLSWDLVGRNLFAMAVEGVVFFLITVLIQYRFFIRPRP	1880
Qy	2021	MPVSTKPVED-DVDVASERQVRVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVR	2079
Db	1881	VNAKLSPLNDEDEDVRRERQRILDGGGQNDILEIKELTKIYRRK---RKPAVDRICVGIP	1937
Qy	2080	PGEFCGLLGVNAGAKTSTFKMLTGDESTTGGEAFVNGHSLKELLQVQQSLGYCPQCDAL	2139
Db	1938	PGEFCGLLGVNAGAKSSTFKMLTGDTTVTRGDAFLNKNLSILNIHEVHQNMGYCPQFDAI	1997
Qy	2140	FDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAI	2199

Db 1998 TELLTGREHVEFFALLRGVPEKEVGKVGEWAIRKLGVLVKYGEKYAGNYSGGNKRKLSTAM 2057

Qy 2200 ALIGYPAFIFLDEPTTGMDPKARRFLWNLIIDLIKTGRSVVLTSHSMEECEALCTRLAIM 2259  
 |||| | :|||||||||||||||| | :| |||||||||||||||||:|||

Db 2058 ALIGGPPVFLDEPTTGMDPKARRFLWNCALSVVKEGRSVVLTSHSMEECEALCTRMAIM 2117

Qy 2260 VNGRLRCLGSIQHLLKNRFGDGYMITVR-TKSSQSVKDVVRFFNRFPEAMLERHHTKVQ 2318  
 |||| ||||:|||||||| | || | : : | | || || :|||:| :|

Db 2118 VNGRFRCLGSVQHLLKNRFGDGYTIVVRIAGSNPDLKPVQDFFGLAFPGSVLKEKHRNMLQ 2177

Qy 2319 YQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFVNFAKKQSDN 2368  
 ||| | |||:| | : | | ||||||||| ||||| |||:

Db 2178 YQLPSSLSSLARIFSILSQSKRLHIEDYSVSQTTLDQVFNFAKDQSD 2227

RESULT 11

US-10-452-510-1

; Sequence 1, Application US/10452510  
 ; Publication No. US20040005666A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hayden, Michael R.  
 ; APPLICANT: Brooks-Wilson, Angela R.  
 ; TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING CHOLESTEROL LEVELS  
 ; FILE REFERENCE: 760050-93  
 ; CURRENT APPLICATION NUMBER: US/10/452,510  
 ; CURRENT FILING DATE: 2003-06-02  
 ; PRIOR APPLICATION NUMBER: US 09/526,193  
 ; PRIOR FILING DATE: 2000-03-15  
 ; PRIOR APPLICATION NUMBER: 60/124,702  
 ; PRIOR FILING DATE: 1999-03-15  
 ; PRIOR APPLICATION NUMBER: 60/138,048  
 ; PRIOR FILING DATE: 1999-06-08  
 ; PRIOR APPLICATION NUMBER: 60/139,600  
 ; PRIOR FILING DATE: 1999-06-17  
 ; PRIOR APPLICATION NUMBER: 60/151,977  
 ; PRIOR FILING DATE: 1999-09-01  
 ; NUMBER OF SEQ ID NOS: 287  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 1  
 ; LENGTH: 2261  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-452-510-1

Query Match 33.5%; Score 4240.5; DB 15; Length 2261;  
 Best Local Similarity 39.8%; Pred. No. 0;  
 Matches 1000; Conservative 345; Mismatches 730; Indels 435; Gaps 61;

Qy 6 QLQLLLWKNVTLKRRSPWVLAFEIFIPVLVFFILLGLRQKKPTISVKEVPFYTAAPL TSA 65  
 ||:|||||:| :|| | | : | ||: | | | : ||

Db 6 QLRLLLWKNLTFRRRQTCQLLLEVAWPLFIFLILISVRLSYPPYEQHECHFNPKA-MPSA 64

Qy 66 GILPVMQSLCPDGQRDEFGFL-----QYANSTVTQLLERLDRVVEEGNLFDPARP 115  
 | || :| : : | : : ||:| | :| :

Db 65 GTLPVWQGIICNANNPCFRYPPTGEAPGVVGNFNKSIVARLFSDARRLL----LYSQKDT 120

Qy 116 SLGSELEALR--QHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNL SLP 173

Db	121	SMKDMRKVLRTLQQIKKSSSNLKLQDFLVDNETFSG-----FLYHNLSLP	165
Qy	174	NSTAQALLAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPA	233
Db	166	KSTVDKMLRADV----ILHKVFLQGYQLHLTS-LCNGS-----KSEEMI-----	204
Qy	234	LLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAELRNQLDV	293
Db	205	---QL---GDQEVSELGCLPREKLAAAE-----RVLRSNMDI	235
Qy	294	AK-VSQQLGLDAPNGSDSSPQAPPPRRLQALLGDLDD-----AQKVLQDQDVLS	341
Db	236	LKPILRTLNSTSPFFSKELAEA--TKTLLHSLGTLAQELFSMRSWSDMRQEVMTLVNS	293
Qy	342	ALALLLPQGACTGRTPGPASGAGGAAN-----GTGAGAVMGPNATAEEGAPSAAALATP	396
Db	294	SSSTQIYQAVSRIVCGHPEGGGLKIKSLNWDYEDNNYKALFGGNGTEEDAETFYDNSTTP	353
Qy	397	---DTLQGQCSAFVQ--LWAGLQPILCGNNRTIEPEALRRGNMSSLGFTSKEQRNLGLLV	451
Db	354	YCNDLMKNLESSPLSRIIWKALKPLLVG-----	381
Qy	452	HLMTSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVWLNISAEIRSFLQGRLLQ	511
Db	382	-----KILYTPDTPATROVMAEVNKTQELAVFHDLEGMWEELSPKIWTFMENSQEMD	434
Qy	512	HLRWL-----QQYVAELRLHPE---ALNLSLDELPPALRQDNFS	547
Db	435	LVRMLLDSRDNDHFWEQQLDGLDWTQAQDIVAFLAKHPEDVQSSNGSVYTWREAFNETN--	492
Qy	548	LPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQDNVTVFAS	607
Db	493	-----QAIRTIS-----RFMECVNLNKLEPIATEVWLINKSME--LLDERKFWAG	535
Qy	608	VIFQTRKDGS--LPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTG---GRFYFLYGFVW	662
Db	536	IVFTGITPGSIELPHHVKYKIRMDIDNVERTNKKIDGYWDPGPRADPFEDMRYVWGGFAY	595
Qy	663	IQDMMERAIIDTFVGHDVVEPGSYVQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSVAM	722
Db	596	LQDVVEQAIIRVLTGTE-KKTGVYMQQMPYPCYVDDIFLRVMSRSMPLFMTLAWIYSVAV	654
Qy	723	TIQHIVAEKEHRLKEVMKTMGLNNAVHWVAFITGFVQLSISVTALTALIKYGQVLMHSH	782
Db	655	IIKGIVYEKEARLKETMRIMGLDNSILWFSWFISSLIPLLVSAGLLVVILKLGNNLLPYSD	714
Qy	783	VVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIYFLSYVPYMYVAIREEVAHDK	842
Db	715	PSVVFVFLSVFAVVTILQCFLISTLFSRANLAAACGGIYFTLYLPYVLC-----VAWQD	769
Qy	843	ITAFE-KCIASLMSTTAFGLGSKYFALYEAVGVGIQWHTFSQSPVEGDDFNLLAVTMLM	901
Db	770	YVGFTLKIFASLLSPVAFGFGCEYFALFEEQGIGVQWQDNLFESPVEEDGFNLTTSVSMML	829
Qy	902	VDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSVM	961



Db 830 FDTFLYGVMTWYIEAVFPGQYGI PRPWYFPCTKSYWFGE---ESDEKSHPGSNQKRIS-- 884  
 Qy 962 EEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLT KVYKDDKKLALNKL SLNLYENQVV 1021  
 : | | | | | | | | : | | | : | : | : | : | :  
 Db 885 ---EIC-----MEEEPHTLKLGVSIQNLVKVYRDGMKVAVDGLALNFYEGQIT 929  
 Qy 1022 SFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLT 1081  
 | | | | | | | | | | | | | | : | | | | | : | | : | | : | | | | | | | |  
 Db 930 SFLGHNGAGKTTTMSILTGLFPPTSGTAYILGKDIRSEMSTIRQNLGVCPQHNVLFDMLT 989  
 Qy 1082 VEEHLWFYSRLKSMAQEEIRREMDKMIEDLEL-SNKRHSLVQTLSSGGMKRKLSVAIAFVG 1140  
 | | | : | | : | | : : | : | : | : | : | | | : | | | : | | | : | | |  
 Db 990 VEEHIWFYARLKGLSEKHVKAEMEQMALDVLGPSSKLKSKTSQLSGGMQRKLSVALAFVG 1049  
 Qy 1141 GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTIILLSTHHMDEADLLGDRIAIISHGKL 1200  
 | : : | | | | | | | | : | | : | : | : | | : | | | | | | | : | | | | | | | |  
 Db 1050 GSKVVILDEPTAGVDPYSRRGIWELLKYRQGRIT ILLSTHHMDEADV LGDRIAIISHGKL 1109  
 Qy 1201 KCCGSPLFLKGT YDGYRLTLVKRPAEPG-----GPQEPGLAS 1238  
 | | | | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 1110 CCVGSSFLKNQLGTGYLTLVKKDVESSLSSCRNSSSTVSYLKKEDSVSQSSSDAGIGS 1169  
 Qy 1239 SPPGRAPLSSCSELQVSQFIRKHVASCLLVSDTSTELSYILPSEAAKKGA FERLFQHLER 1298  
 | : | | | | : | | | | | : | | : | : | | | : | | | : | | : | :  
 Db 1170 DHESDTLTIDVS--AISNLIRKHVSEARLVEDIGHELT YVLPYEA AKEGAFVELFHEIDD 1227  
 Qy 1299 SLDALHLSSFGLMDTTLEEVLKVSEEDQSLENSEADV KESRKDVLPGAEGPASGEGHAG 1358  
 | | : | : : | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :  
 Db 1228 RLSDLGISSYGISETTLEE IFLKVAEE-----SGVDA-ETSDGTLP----- 1267  
 Qy 1359 NLARCELTQSQASLQSASSVGSARGDEGAGYTDVYGDYRPLF-DNPQDPD--NVS LQEV 1415  
 | | : | : | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :  
 Db 1268 -----ARRNRRA-FGDKQ SCLRPFTEDDAADPNDSIDIPESR 1303  
 Qy 1416 EAEALSRV-GQGSRKLDGGWLKVRFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMT 1474  
 | : | | : | : | : | : | : | | | | | | | | : | | : | | : | | : | | : | | : | | : | | :  
 Db 1304 ETDLLSGMDGKGSYQVKGWKLTQQQFVALLWKRLLIARRSRKGFFAQIVLPAVFVCI ALV 1363  
 Qy 1475 VALSVPEIGDLPLPLVLSPSQYH-NYTQPRGNFI PYANEERREYRLRLSPDASPQQLVSTF 1533  
 : | | | | | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :  
 Db 1364 FSLIVPPFGKYPSLELQPW MYNEQYT-----FVSNDAP E-----DTGTLELLNAL 1408  
 Qy 1534 RLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPP 1593  
 | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 1409 TKDPGF GTRCM-----EGNPI----- 1424  
 Qy 1594 PSPAPSDSPASPD EDLQAWNVS LPPTAGPEMWT SAPSLPRLVREPVR-----C 1641  
 | | | | | | | | | | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 1425 -----PD-----TPCQAGEEWT TAP-VPQTIMDLFQNGNWTM QNPSPAC 1463  
 Qy 1642 TCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILT DITGHN VSEYLLFTSDRF----- 1690  
 | | : | | | | | | | : | | | | : | | : | | : | | : | | : | | : | | : | | : | | :  
 Db 1464 QCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKN 1523  
 Qy 1691 ----RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK----- 1721  
 | | | : | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 1524 KIWVNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTG 1583

Qy 1722 IAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSAS 1781  
 : | :||:||||:|:: :|| :||||||| | : ||: |||| |||:| |  
 Db 1584 LDTRNNVKVWFNNKGWHAISSFLNVINNAILRANLQKGE-NPSHYGITAFNHPLNLTKQQ 1642

Qy 1782 LS-LDYLLQGTDVVIAIFIIVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLAN 1840  
 || : : ||::| :| |||||||||||||: | : |||||||:| | :|||:|  
 Db 1643 LSEVALMTTSVDVLVSICVIFAMSFVPASFVFLIQERVSKAKHLQFISGVKPVYWLNS 1702

Qy 1841 YVWDMNLNYPATCCVILFVFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWFEV 1900  
 :||| ||:||| :|| | :| | | | : | |||||||:||||| |::  
 Db 1703 FVWDMCNVVPATLVIIIFICFQQKSYVSSSTNLPVLALLLLLYGWSITPLMPASFVFKI 1762

Qy 1901 PSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVVNSYLKSCFLIFPNYNLGHGLME 1960  
 ||:|| | :||||| :||||:| | | :| || | |||:| | ||:|  
 Db 1763 PSTAYVVLTSVNLFIGINGSVATFVLELEFT-DNKLNNDILKSVFLIFPHFCLGRGLID 1821

Qy 1961 MAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQR 2020  
 | | : : : | : : || ||:| | | ||||||| ||:|:| | | ||:  
 Db 1822 MVKNQAMADALERFGE-NRFVSPLSWDLVGRNLFAMAVEGVVFFLITVLIQYRFFIRPRP 1880

Qy 2021 MPVSTKPVED-DVDVASERQVRVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVR 2079  
 : | : | | | ||||:| | ||::|: |||:| : | |||:|:|  
 Db 1881 VNAKLSPLNDEDEDVRRERQRILDGGGQNDILEIKELTKIYRRK---RKPAVDRICVGIP 1937

Qy 2080 PGECFGLLVNGAGKTSTFKMLTGDESTTGGEAFVNGHSLVKELLQVQQSLGYCPCDAL 2139  
 |||||||||||||:||||||| : | |::| :|:| : :| |::||| ||:  
 Db 1938 PGECFGLLVNGAGKSSTFKMLTGDTTVTRGDAFLNKNSILSNIHEVHQNMGYCPCQDAI 1997

Qy 2140 FDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAI 2199  
 : || |||: : |||: | : :| :||: || | || :| || |||||||||:  
 Db 1998 TELLTGREHVEFFALLRGVPEKEVGKVGWAIKRLGLVKYGEKYAGNYSGGNKRKLSTAM 2057

Qy 2200 ALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIM 2259  
 |||| | :||||||||||||||| | :| |||||||||||||||||:|||  
 Db 2058 ALIGGPPVFLDEPTTGMDPKARRFLWNCALSUVKEGRSVVLTSHSMEECEALCTRMAM 2117

Qy 2260 VNGRLRCLGSIQHLKNRFGDGYMITVR-TKSSQSVKDVVRFFNRNFPEAMLERHHTKVQ 2318  
 |||| ||||:||||||| | || | : :| | | || :|||:| :|  
 Db 2118 VNGRFRCLGSVQHLKNRFGDGYTIVVRIAGSNPDLKPVQDFFGLAFPGSVLKEKHRNMLQ 2177

Qy 2319 YQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFVNFAKKQSDN 2368  
 ||| | |||:||:| | | ||||||||| ||||| |||:  
 Db 2178 YQLPSSLSSLARIFSILSQSKRLHIEDYSVSQTTLDQVFNFAKDQSD 2227

RESULT 12

US-10-745-377-5

; Sequence 5, Application US/10745377

; Publication No. US20040137423A1

; GENERAL INFORMATION:

; APPLICANT: Hayden, Michael R.

; APPLICANT: Pimstone, Simon

; APPLICANT: Brooks-Wilson, Angela R.

; APPLICANT: Clee, Susanne M.

; TITLE OF INVENTION: Compositions and Methods for Modulating

; TITLE OF INVENTION: HDL Cholesterol and Triglyceride Levels

```
; FILE REFERENCE: 760050-109
; CURRENT APPLICATION NUMBER: US/10/745,377
; CURRENT FILING DATE: 2003-12-23
; PRIOR APPLICATION NUMBER: 09/654,323
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: US 60/124,702
; PRIOR FILING DATE: 1999-03-15
; PRIOR APPLICATION NUMBER: US 60/138,048
; PRIOR FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: US 60/139,600
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/151,977
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: US 09/526,193
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: US 60/213,958
; PRIOR FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 256
; SOFTWARE: Word for Windows Version 6.0 (ASCII Text)
; SEQ ID NO 5
; LENGTH: 2261
; TYPE: PRT
; ORGANISM: homo sapien
US-10-745-377-5
```

```
Query Match          33.5%; Score 4240.5; DB 16; Length 2261;
Best Local Similarity 39.8%; Pred. No. 0;
Matches 1000; Conservative 345; Mismatches 730; Indels 435; Gaps 61;
```

```
Qy      6 QLQLLLWKNVTLKRRSPWVLAFEIFIPVLVFFILLGLRQKKPTISVKEVPFYTAAPL TSA 65
      ||:|||||:| :|| | |: || :| ||: :| | | | : ||
Db      6 QLRLLLWKNLTFRRRQTCQLLLEVAWPLFIFLILISVRLSYPPYEQHECHFPNKA-MPSA 64

Qy     66 GILPVMQSLCPDQQRDEFGL-----QYANSTVTQLLERLDRVVEEGNLFDPARP 115
      | || :| : : | : : | | :| | :| | :
Db     65 GTLPWWQGIICNANNPCFRYPTPGEAPGVVGNFNKSIVARLFSDARRLL-----LYSQKDT 120

Qy    116 SLGSELEALR--QHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSLP 173
      |: : || | :: |: :| || | | | | || ||||
Db    121 SMKDMRKVLRTLQQIKSSSNLKLQDFLVDNETFSG-----FLYHNLSLP 165

Qy    174 NSTAQALIAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPA 233
      || :| || | :| :| | | | | | : ||::
Db    166 KSTVDKMLRADV----ILHKVFLQGYQLHLTS-LCNGS-----KSEEMI----- 204

Qy    234 LLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAELRNQLDV 293
      || | |: : :| : | : | | | | | | | | | | | | | | | | | | | | | |
Db    205 ---QL-----GDQEVSELCGLPREKLAAAE-----RVLRSNMDI 235

Qy    294 AK-VSQQLGLDAPNGSDSSPQAPPPRRLQALLGDLLD-----AQKVLQDQDVLS 341
      | : : | :| | :| :| | | | | | | | | | | | | | | | | | | | | | |
Db    236 LKPILRTLNSTSPFFPSKELAEA--TKTLLHSLGTLAQELFSMRSWSDMRQEVMTLVNVS 293

Qy    342 ALALLLPQGACTGRTPGPASGAGGAAN-----GTGAGAVMGPNATAEEGAPSAAALATP 396
      : : | : | | | : | : | | | | | | | | | | | | | | | | | | | | | |
Db    294 SSSSTQIYQAVSRIVCGHPEGGGLKIKSLNWWYEDNNYKALFGGNGTEEDAETFYDNSTTP 353
```

Qy 397 ---DTLQGQCSAFVQ--LWAGLQPILCGNNRTIEPEALRRGNMSSLGFTSKEQRNLGLLV 451  
| : : | : : | : | : | : | : |

Db 354 YCNDLMKNLESSPLSRIIWKALKPLLVG----- 381

Qy 452 HLM/TSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVWLNISAEIRSFLEQGRLOO 511  
| | | | | : | : : | : | : : | : | : | : | : |

Db 382 -----KILYTPDTPATROVMAEVNKTQELAVFHDLEGMWEELSPKIWTFMENSQEMD 434

Qy 512 HLRWL-----QQYVAELRLHPE---ALNLSLDELPPALRQDNFS 547  
: | | | : | | | | | : | | : | : |

Db 435 LVRMLLDSRDNDHFWEQQLDGLDWTQAQDIVAFLAKHPEDVQSSNGSVYTWAREAFNETN-- 492

Qy 548 LPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPEESIVNYTLNQAYQDNVTVFAS 607  
| : | | : | | : | : : | : | : | : | : |

Db 493 -----QAIRTIS-----RFMECVNLNKLEPIATEVWLINKSME--LLDERKFWAG 535

Qy 608 VIFQTRKDGS--LPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTG---GRFYFLYGFVW 662  
: : | | | | | | : : | : | : | | | | | : | : |

Db 536 IVFTGITPGSIELPHHVKYKIRMDIDNVERTNKKIKDGYWDPGPRADPFEDMRYVWGGFAY 595

Qy 663 IQDMMERAIIDTFVGHADVPEPGSYVQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSVAM 722  
: | : : : | : | : : | : | : | : | : | : | : | : | : | : |

Db 596 LQDVVEQAIIRVLTGTE-KKTGVYMQQMPYPCYVDDIFLRVMSRSMPLFMTLAWIYSVAV 654

Qy 723 TIQHIVAEKEHRLKEVMKTMGLNNAVHWVAFITGFVQLSISVTALTALIKYGQVLMHSH 782  
| : | | | | | : | : | : | : | : | : | : | : | : | : |

Db 655 IIKGIVYEKEARLKETMRIMGLDNSILWFSWFISSLIPLLVSAGLLVVLKLGNNLPYSD 714

Qy 783 VVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHDK 842  
: : : : | : | : | : | : | : | : | : | : | : | : |

Db 715 PSVVFVFLSVFAVVTILQCFLISTLFSRANLAAACGGIIYFTLYLPYVLC-----VAWQD 769

Qy 843 ITAFE-KCIASLMSTTAFGLGSKYFALYEAVGVGIQWHTFSQSPVEGDDFNLLAVTMLM 901  
| | | : | | | : | : | : | : | : | : | : | : | : | : |

Db 770 YVGFTLKIFASLLSPVAFGFGCEYFALFEEQGIGVQWDLNLFESPVEEDGFNLTTSVSMML 829

Qy 902 VDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPPWARTPRLSVM 961  
| : | : : | : | : | : | : | : | : | : | : | : |

Db 830 FDTFLYGVMTWYIEAVFPQGQYGI PRPWYFPCTKSYWFGE---ESDEKSHPGSNQKRIS-- 884

Qy 962 EEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTQVYKDDKKLALNKLSLNLYENQVV 1021  
: | | | | | | | | : | : | : | : | : | : | : |

Db 885 ---EIC-----MEEEPHTLKLGVSIQNLVKVYRDGMKVAVDGLALNFYEGQIT 929

Qy 1022 SFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLT 1081  
| | | | | | | | | | | | | | : | | | : | : | : | : | : | : |

Db 930 SFLGHNGAGKTTTMSILTGLFPPTSGTAYILGKDIRSEMSTIRQNLGVCPQHNVLFDMLT 989

Qy 1082 VEEHLWFYSRLKSMAQEEIRREMDKMIEDLEL-SNKRHSLVQTLSSGGMKRKLSVAIAFVG 1140  
| | | : | : | : | : : : | : | : | : | : | : | : | : |

Db 990 VEEHIWFYARLKGLSEKHVKAEMEQLMALDVLGPSSKLSKTSQSLSGGMQRKLSVALAFVG 1049

Qy 1141 GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTIILLSTHMDLGLDRIAIISHGKL 1200  
| : : | : | : | : | : | : | : | : | : | : | : | : | : |

Db 1050 GSKVVILDEPTAGVDPYSRRGIWELLKYRQGRTIILLSTHMDLGLDRIAIISHGKL 1109

Qy 1201 KCCGSPLFLKGTGYGDGYRLTLVKRPAEPG-----GPQEPGLAS 1238

Db	1110	CCVGSSIFLKNQLGTGYLTLVKKDVESSLSSCRNSSSTVSYLKKEDSVSQSSSDAGLGS	1169
Qy	1239	SPPGRAPLSSCSELQVSQFIRKHAVASCLLVSDTSTELSYILPSEAAKKAFAERLFQHLE	1298
Db	1170	DHESDTLTIDVS--AISNLIIRKHVSEARLVEDIGHELTYVLPYEAKEGAFVELFHEIDD	1227
Qy	1299	SLDALHLSSFGLMDTTLEEVFLKVSEEDQSLENSEADVKESRKDVLPGAEGPASGEGHAG	1358
Db	1228	RLSDLGISYGISETTLEEIFLKVAEE-----SGVDA-ETSDGTLP-----	1267
Qy	1359	NLARCELTQSQASLQSAASSVGSARGDEGAGYTDVYGDYRPLF-DNPQDPD--NVS	1415
Db	1268	-----ARRNRA-FGDKQSCLRPFTEDDAADPNDSIDIPESR	1303
Qy	1416	EAEALSRV-GQGSRKLDGGWLKVRQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMT	1474
Db	1304	ETDLLSGMDGKGSYQVKGWKLTTQQQFVALLWKRLLIARRSRKGFFAQIVLPAVFCIALV	1363
Qy	1475	VALSVPEIGDLPPLVLSPSQYH-NYTQPRGNFIPIYANEERREYRLRLSPDASPQQLVSTF	1533
Db	1364	FSLIVPPFGKYPSELELQPMWYNEQYT-----FVSNDAP-----DTGTLELLNAL	1408
Qy	1534	RLPSGVGATCVLKSPANGSLGPTLNLSGSRLLAARFFDSMCLESFTQGLPLSNFVPPP	1593
Db	1409	TKDPGFGTRCM-----EGNFI-----	1424
Qy	1594	PSPAPSDSPASPDEDLQAWNVS LPPTAGPEMWTSA PSLPRLVREPVR-----C	1641
Db	1425	-----PD-----TPCQAGEEWTAP-VPQTIMDLFQNGNWTM QNPSPAC	1463
Qy	1642	TCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILTDITGHNVSEYLLFTSDRF-----	1690
Db	1464	QCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKN	1523
Qy	1691	----RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK-----	1721
Db	1524	KIWNNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKHKLAKDSSADRFLNSLGRFMTG	1583
Qy	1722	I AVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSAS	1781
Db	1584	LDTRNNVKVWFNNKGWHAISSFLNVINNAILRANLQKGE-NPSHYGITAFNHPNLTKQQ	1642
Qy	1782	LS-LDYLLQGTDVVIAIFIIVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLAN	1840
Db	1643	LSEVALMTTSVDVLVSICVIFAMSFVPASFVFLIQERVSKAKHLQFISGVKPVYIYWSN	1702
Qy	1841	YVWDMNLNYLVPATCCVILFVFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWFEV	1900
Db	1703	FVWDMCNVVPATLVIIIFICFQQKSYVSSTNLPVLALLLLLYGWSITPLMYPASFVFKI	1762
Qy	1901	PSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSYKSCFLIFPNYNLGHGLME	1960
Db	1763	PSTAYVVLTSVNLFINGSVATFVLELFT-DNKLNNINDILKSVFLIFPHFCLGRGLID	1821
Qy	1961	MAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQR	2020

Db 1822 MVKNQAMADALERFGE-NRFVSPLSWDLVGRNLFAMAVEGVVFFLITVLIQYRFFIRPRP 1880

Qy 2021 MPVSTKPVED-DVDVASERQRLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVR 2079  
: | : | | | | | | | | | | : | | : : : | | : : : | | : | : | :

Db 1881 VNAKLSPLNDEDEDVRRERQRILDGGGQNDILEIKELTKIYRRK---RKPAVDRICVGIP 1937

Qy 2080 PGEFCGLLGVNGAGKTSTFKMLTGDESTTGGEAFVNGHSLKELLQVQQSLGYCPQCDAL 2139  
| | | | | | | | | | : | | | | | | : | | : | : | : | : | : | : | | :

Db 1938 PGEFCGLLGVNGAGKSSTFKMLTGDTTVTRGDAFLNKNLSILNIHEVHQNMGYCPQFDAI 1997

Qy 2140 FDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAI 2199  
: | | | : : | : | : | : | : | : | : | : | : | : | : | : | : | :

Db 1998 TELLTGREHVEFFALLRGVPEKEVGKVGEWAIRKGLVKYGEKYAGNYSGGNKRKLSTAM 2057

Qy 2200 ALIGYPAFIFLDEPTTGMDPKARRFLWNLIIDLIKTGRSVVLTSHSMEECEALCTRLAIM 2259  
| | | | : | | | | | | | | | | | | : | | | | | | | | | | : | | | :

Db 2058 ALIGGPPVVFLDEPTTGMDPKARRFLWNCALSVVKEGRSVVLTSHSMEECEALCTRMAM 2117

Qy 2260 VNGRLRCLGSIQHLKNRFGDGYMITVR-TKSSQSVKDVVREFNRNFPPEAMLKERHHTKVQ 2318  
| | | | | | : | | | | | | | | | | : | : | | | | | | : | : | : | :

Db 2118 VNGRFRCLGSVQHLKNRFGDGYTIVVRIAGSNPDLKPVQDFFGLAFPGSVLKEKHRNMLQ 2177

Qy 2319 YQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFVNFAKKQSDN 2368  
| | | | | : : | : | | | | | | | | | | | | : | : | : | :

Db 2178 YQLPSSLSSLARIFSILSQSKRLHIEDYSVSQTTLDQVFNFAKDQSD 2227

RESULT 13

US-10-744-465-1

; Sequence 1, Application US/10744465  
; Publication No. US20040157250A1  
; GENERAL INFORMATION:  
; APPLICANT: Hayden, Michael R.  
; APPLICANT: Brooks-Wilson, Angela R.  
; APPLICANT: Pimstone, Simon N.  
; TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING CHOLESTEROL LEVELS  
; FILE REFERENCE: 760050-92  
; CURRENT APPLICATION NUMBER: US/10/744,465  
; CURRENT FILING DATE: 2003-12-23  
; PRIOR APPLICATION NUMBER: 10/617,334  
; PRIOR FILING DATE: 2003-07-10  
; PRIOR APPLICATION NUMBER: US 09/526,193  
; PRIOR FILING DATE: 2000-03-15  
; PRIOR APPLICATION NUMBER: 60/124,702  
; PRIOR FILING DATE: 1999-03-15  
; PRIOR APPLICATION NUMBER: 60/138,048  
; PRIOR FILING DATE: 1999-06-08  
; PRIOR APPLICATION NUMBER: 60/139,600  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: 60/151,977  
; PRIOR FILING DATE: 1999-09-01  
; NUMBER OF SEQ ID NOS: 287  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2261  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-10-744-465-1

Query Match 33.5%; Score 4240.5; DB 16; Length 2261;  
Best Local Similarity 39.8%; Pred. No. 0;  
Matches 1000; Conservative 345; Mismatches 730; Indels 435; Gaps 61;

```
Qy      6 QLQLLLWKNVTLKRRSPWVLAFEIFIPVLVFFILLGLRQKKPTISVKEVPFYTAAPL TSA 65
      ||:|||||:| :|| | | : || :| ||: :| | | | : ||
Db      6 QLRLLLWKNLTFRRRQTCQLLLEVAWPLFIFILILISVRLSYPPYEQHECHFPNKA-MPSA 64

Qy     66 GILPVMQSLCPDGQRDEFGFL-----QYANSTVTQLLERLDRVVEEGNLFDPARP 115
      | || :| : : | : | : | | :| | :| | :| :|
Db     65 GTLPWVQGIICNANNPCFRYPTPGEAPGVVGNFNKSIVARLFSDARRLL----LYSQKDT 120

Qy    116 SLGSELEALR--QHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSLP 173
      |: : || | : : |: | : | | | | | | | | | | | |
Db    121 SMKDMRKVLRTLQQIKKSSSNLKLQDFLVDNETFSG-----FLYHNLSLP 165

Qy    174 NSTAQALLAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPA 233
      || :| | | : | :| | | | | | | | | | : ||:
Db    166 KSTVDKMLRADV----ILHKVFLQGYQLHLTS-LCNGS-----KSEEMI----- 204

Qy    234 LLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAELRNQLDV 293
      || | | | : : :| : | : | | | | | | | | | |
Db    205 ---QL---GDQEVSELCLPREKLAAAE-----RVLRSNMDI 235

Qy    294 AK-VSQQGLDAPNGSDSSPQAPPPRRLQALLGLDLD-----AQKVLQDQDVVLS 341
      | : : | :| | :| :| :| | | | | | | | | | | |
Db    236 LKPILRTLNSTSPFPSKELAEA--TKTLHSLGTLAQELFSMRSWSDMRQEVMTNVNS 293

Qy    342 ALALLLPQGACTGRTPGPPASGAGGAAN-----GTGAGAVMGNATAEEGAPSAAALATP 396
      : : | : | | | : | : | | | | : | | | | |
Db    294 SSSSTQIYQAVSRIVCGHPEGGLKIKSLNWWYEDNNYKALFGNGTEEDAETFYDNSTTP 353

Qy    397 ---DTLQGQCSAFVQ--LWAGLQPILCGNNRTIEPEALRRGNMSSLGFTSKEQRNLGLLV 451
      | : : | : : :| | :| :| | |
Db    354 YCNLDMKNLESSPLSRIWKALKPLLVG----- 381

Qy    452 HLMTSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVWLNISAEIRSFLEQGRLOQ 511
      |||| | :| : : | :| : | :| :| :| :| :| :|
Db    382 -----KILYTPDTPATROVMAEVNKTFOELAVFHDLEGMWEELSPKIWTFMENSQEMD 434

Qy    512 HLRWL-----QQYVAELRLHPE---ALNLSLDELPPALRQDNFS 547
      :| | | | | | | | | | | | | | | | | | | |
Db    435 LVRMLLDSRDNDHFEWQQLDGLDWT AQDIVAFLAKHPEDVQSSNGSVYTWREAFNETN-- 492

Qy    548 LPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPEESIVNYTLNQAYQDNVTVFAS 607
      | : || :|| | : : : | :| :| :| :| :|
Db    493 -----QAIRTIS-----RFMECVNLNKLEPIATEVWLINKSME--LLDERKFWAG 535

Qy    608 VIFQTRKDG--LPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTG---GRFYFLYGFVW 662
      : :| | | | | | | | : | :| :| :| | | | | | |
Db    536 IVFTGITPGSIELPHHVYKIRMDIDNVERTNKIKDGYWDPGPRADPFEDMRYVWGGFAY 595

Qy    663 IQDMMERAIIDTFVGHDVVEPGSYVQMFYPYCYTRDDFLFVIEHMMPLCMVISWVYSVAM 722
      :||: :| :|| | : : | | :| | | | | | | | :| :| :| :| :|
Db    596 LQDVVEQAIIRVLTGTE-KKTGVYMQMPYPCYVDDIFLRVMSRSMPLFMTLAWIYSVAV 654
```

Qy 723 TIQHIVAEKEHRLKEVMKTMGLNNAVHWVAWFITGEVQLSISVTALTALILKYQVLMHSH 782  
 | : || ||| |||| | : ||| : : | : | | ||| | : : |  
 Db 655 IIKGIVYEKEARLKETMRIMGLDNSILWFSWFISSLIPLLVSAGLLVVILKLGNNLLPYSD 714  
 Qy 783 VVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHDK 842  
 : : : ||| : || ||| : | : | : ||| ||||| | : || : ||  
 Db 715 PSVVFVFLSVFAVVTILQCFLISTLFSRANLAAACGGIIYFTLYLPYVLC-----VAWQD 769  
 Qy 843 ITAFE-KCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTMLM 901  
 | | ||| : | || | : ||| : | : | : | : ||| | ||| : | : :  
 Db 770 YVGFTLKFASLLSPVAFGFGCEYFALFEEQGIGVQWDNLFESPVEEDGFNLTTSVSMML 829  
 Qy 902 VDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSVM 961  
 | : ||| : ||||| || ||| : ||||| |||| | : | | | : | : |  
 Db 830 FDTFLYGVMTWYIEAVFPQYGI PRPWYFPCTKSYWFG-----ESDEKSHPGSNQKRIS-- 884  
 Qy 962 EEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTKVYKDDKKLALNKLNLNLYENQVV 1021  
 : | ||||| | | : | ||| : | : | : | : || | | :  
 Db 885 ---EIC-----MEEEPHTLKLGVSIQNLVKVYRDGMKVAVDGLALNFYEGQIT 929  
 Qy 1022 SFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRIT 1081  
 ||||| ||||| ||||| ||||| : | | ||| : || : ||| : ||||| ||  
 Db 930 SFLGHNGAGKTTTMSILTGLFPPTSGTAYILGKDIRSEMSTIRQNLGVCQHNVLFDMLT 989  
 Qy 1082 VEEHLWFYSRLKSMAQEEIRREMDKMIEDLEL-SNKRHSLVQTLSSGGMKRKLSVAIAFVG 1140  
 |||| : ||| : || : : : || : | : | | | |||| : ||||| : ||||  
 Db 990 VEEHIWFYARLKLSEKHVKAEMEQLMALDVGLPSSKLKSKTSQLSGGMQRKLSVALAFVG 1049  
 Qy 1141 GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL 1200  
 || : ||||| ||||| : || ||| : ||| : |||| : ||||| ||||| : ||||| |||||  
 Db 1050 GSKVVILDEPTAGVDPYSRRGIWELLLKYRQGRTIILSTHHMDEADVLGDRIAIISHGKL 1109  
 Qy 1201 KCCGSPLFLKGTGTYGDRYRLTLVKRPAEPG-----GPQEPGLAS 1238  
 | || |||| | || |||| : | : || | : || | : || |  
 Db 1110 CCVGSSFLKNQLGTGYLLTLVKKDVESLSSCRNSSSTVSYLKKEDSVSQSSSDAGLGS 1169  
 Qy 1239 SPPGRAPLSSCSELQVSQFIRKHVASCLLVSDTSTELSYILPSEAAKGAERLRFQHLER 1298  
 | : | |||| : || | ||| : || |||| : || || : :  
 Db 1170 DHESDTLTIDVS--AISNLIRKHVSEARLVEDIGHELTIVLPYEAKEGAFVELFHEIDD 1227  
 Qy 1299 SLDALHLSSFGIMDTTLEEVLKVSEEDQSLENSEADVKE SRKDVLPGAEGPASGEGHAG 1358  
 | | : ||| : : |||| : |||| : || | | | : ||  
 Db 1228 RLSDLGISSYGISETTLEEIFLKVAEE-----SGVDA-ETSDGTLP----- 1267  
 Qy 1359 NLARCSELTQSQASLQSASSVGSARGDEGAGYTDVYGDYRPLF-DNPQDPD--NVSILQEV 1415  
 || : | : | || | : || : : :  
 Db 1268 -----ARRNRA-FGDKQSCLRPFTEDDAADPNDSIDDPESR 1303  
 Qy 1416 EAEALSRV-GQGSRKLDGGWLKVRQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMT 1474  
 | : || : | : || : : | | : || || || |||| : | | : || : || |||| : | :  
 Db 1304 ETDLLSGMDGKGSYQVKGWKLTTQQQFVALLWKRLLIARRSRKGFFAQIVLPAVFCIALV 1363  
 Qy 1475 VALSVPEIGDLPLVLSPSQYH-NYTQPRGNFIPYANEERREYRLRLSPDASPQQVLVSTF 1533  
 : | || | | | | | : || : : : | | : : :  
 Db 1364 FSLIVPPFGKYPSLELQPMWYNEQYT-----FVSNDAP-----DTGTLELLNAL 1408



Qy	1534	RLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPP	1593
Db	1409	TKDPGFGTRCM-----EGNPI-----	1424
Qy	1594	PSPAPSDSPASPDDELQAWNVS LPPTAGPEMWTSA PSLPRLVREPVR-----C	1641
Db	1425	-----PD-----TPCQAGEEEWTTAP-VPQTIMDLFQNGNWTM QNPSPAC	1463
Qy	1642	TCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILT DITGHNVS EYLLFTSDRF-----	1690
Db	1464	QCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKN	1523
Qy	1691	----RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK-----	1721
Db	1524	KIWNNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTG	1583
Qy	1722	IAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHMPMNKTSAS	1781
Db	1584	LDTRNNVVKWFENNKGWHAISSFLNVINNA ILRANLQKGE-NPSHYGITAFNHPLNLTKQQ	1642
Qy	1782	LS-LDYLLQGTDVVIAIFIIVAMS FVPASFVVFLVAEKSTKAKHLQFVSGCNP I IYWLAN	1840
Db	1643	LSEVALMTTSDVDLVLSICVIFAMSFVPASFVVFLIQERVSKAKHLQFISGVKPIYWLNSN	1702
Qy	1841	YVWDMNLNYLVPATCCV I I L F V F D L P A Y T S P T N F P A V L S L F L L Y G W S I T P I M Y P A S F W F E V	1900
Db	1703	FVWDMCNYVVPATLVIIIFICFQQKSYVSSTNL PVLALLLLLYGWSITPLMPYAS FVFKI	1762
Qy	1901	PSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLME	1960
Db	1763	PSTAYVVLTSVNLFIGINGSVATFVLELFT-DNKLNNINDILKSVFLIFPHFCLGRGLID	1821
Qy	1961	MAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEG VVG FLLTIMCQYNFLRRPQR	2020
Db	1822	MVKNQAMADALERFGE-NRFVSPLSWDLVGRNLFAMAVEG VVFLLITVLIQYRFFIRPRP	1880
Qy	2021	MPVSTKPVED-DVDVASERQVR LRGDADNDMVKIENLTKVYKSRKIGRILAVDR LCLGVR	2079
Db	1881	VNAKLSPLNDEDEDVRRERQRILDGGGQNDILEIKELTKIYRRK---RKPAVDRICVGIP	1937
Qy	2080	PGEFCGLLG VNGAGKTSTFKMLTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDAL	2139
Db	1938	PGEFCGLLG VNGAGKSSTFKMLTGDTTVTRGDAFLNKNSILSNIHEVHQNMGYCPQFDAI	1997
Qy	2140	FDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELT KYADKPAGTYSGGNKRKLSTAI	2199
Db	1998	TELLTGREHVEFFALLRGVPEKEVGKVG EWAIRKLG LVKYGEKYAGNYSGGNKRKLSTAM	2057
Qy	2200	ALIGYPAFIFLDEPTTGMDPKARRFLWNLI LDIKTGRSVVLTSHSMEECEALCTR LAIM	2259
Db	2058	ALIGGPPVVF LDEPTTGMDPKARRFLWNCALSVVKEGRSVVLTSHSMEECEALCTRMAIM	2117
Qy	2260	VNGRLRCLGSIQHLKNRFGDGYMITVR-TKSSQSVKDVVRFFNRNFPEAMLKERHHTKVQ	2318
Db	2118	VNGRFRCLG SVQHLKNRFGDGYTIVVRIAGSNPDLKP VQDFFGLAFPGSVLKEKHRNMLQ	2177
Qy	2319	YQLKSEHISLAQVFSKMEOVSGVLGIEDYSVSQTTLDNVFVNFAKKQSDN	2368



Qy 342 ALALLLPQGACTGRTPGPPASGAGGAAN-----GTGAGAVMGPNATAEEGAPSAAALATP 396  
: : | : | | : : | : | | | : | |

Db 294 SSSSTQIYQAVSRIVCGHPEGGLKIKSLNWDYEDNNYKALFGNGTEEDAETFFYDNSTTP 353

Qy 397 ---DTLQGCQSAFVQ--LWAGLQPILCGNNRTIEPEALRRGNMSSLGFTSKEQRNLGLLV 451  
| : : | : : : | | : | : |

Db 354 YCNDLMKNLESSPLSRIIWKALKPLLVG----- 381

Qy 452 HLMTSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVWINISAEIRSFEQGRLOQ 511  
| | | | | : | : : | : | : : | : | : | : |

Db 382 -----KILYTPDTPATRQVMAEVNKTQELAVFHDLEGMWEELSPKIWTFMENSQEMD 434

Qy 512 HLRWL-----QQYVAELRLHPE---ALNLSLDELPPALRQDNFS 547  
: | | | | | | | | | | : | : | : |

Db 435 LVRMLLDSRDNDHFWEQQLDGLDWTQADIVAFIAKHPEDVQSSNGSVYTWREAFNETN-- 492

Qy 548 LPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPEESIVNYTLNQAYQDNVTVFAS 607  
| : | | : | | | : : | : : | : |

Db 493 -----QAIRTIS-----RFMECVNLNKLEPIATEVWLINKSME--LLDERKFWAG 535

Qy 608 VIFQTRKDGS--LPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTG---GRFYFLYGFVW 662  
: : | | | | | | : | : | : | | | | | : |

Db 536 IVFTGITPGSIELPHHVYKIRMDIDNVERTNKIKDGYWDPGPRADPFEDMRYVWGGFAY 595

Qy 663 IQDMMERAIIDTFVGHVVEPGSYVQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSVAM 722  
: | : : | : | | : : | : | | | | | | : | : | : | : | : |

Db 596 LQDVVEQAIIRVLTGTE-KKTGVYMQQMPYPCYVDDIFLRVMSRSMPLFMTLAWIYSVAV 654

Qy 723 TIQHIVAKEHRLKEVMKTMGLNNAVHWWAFITGFVQLSISVTALTALIKYQVLMHSH 782  
| : | | | | | | : | : | : | : | : | : | : | : |

Db 655 IIKGIVYEKEARLKETMRIMGLDNSILWFSWFISSLIPLVLSAGLLVILKLGKLNLLPYSD 714

Qy 783 VVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIFLSYVPYMYVAIREEVAHDK 842  
: : : : | : | : | : | : | : | : | : | : | : | : |

Db 715 PSVVVFVFLSVFAVVTILQCFLISTLFSRANLAAACGGIIFLYLPVLC-----VAWQD 769

Qy 843 ITAFE-KCIASIMSTTAFGLGSKYFALYEAVGVGIQWHTFSQSPVEGDDFNLLAVTMLM 901  
| | | : | | | : | : | : | : | : | : | : | : |

Db 770 YVGFTLKIFASLLSPVAFGFGCEYFALFEEQGIGVQWDNLFESPVEEDGFNLTTSVSMML 829

Qy 902 VDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSVM 961  
| : | : : | : | : | : | : | : | : | : | : | : | : |

Db 830 FDTFLYGVMTWYIEAVFPGQYGIIPRPWYFPCTKSYWFGE---ESDEKSHPGSNQKRIS-- 884

Qy 962 EEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTQVYKDDKKLALNKLNLNLYENQVV 1021  
: | | | | | | | | : | : | : | : | : | : |

Db 885 ----EIC-----MEEEPHTLKLGVSIQNLVKVYRDGMKVAVDGLALNFYEGQIT 929

Qy 1022 SFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLT 1081  
| | | | | | | | | | | | : | | | : | | : | : | : | : |

Db 930 SFLGHNGAGKTTTMSILTGLFPPTSGTAYILGKDIRSEMSTIRQNLGVCQPQHNVLFDMLT 989

Qy 1082 VEEHLWFYSRLKSMAQEEIRREMDKMIEDLEL-SNKRHSLVQTLSSGGMKRKLSVAIAFVG 1140  
| | : : | : | : : : : | : | : | | | : | : | : | : |

Db 990 VEEHIWFYARLKGLSEKHVKAEMEQLDVGLPSSKLKSKTSQSLSGGMQRKLSVALAFVG 1049

Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTIILSTHHMDEADLLGDRIAIISHGKL	1200
Db	1050	GSKVVILDEPTAGVDPYSRRGIWELLKRYRQGRITILSTHHMDEADVLGDRIAIISHGKL	1109
Qy	1201	KCCGSPLFLKGTYG DG YRLTLVKRPAEPG-----GPQEPGLAS	1238
Db	1110	CCVGSSSLFLKNQLGTGYLTLVKKDVESSLSSCRNSSSTVSYLKKEDSVSQSSSDAGLGS	1169
Qy	1239	SPPGRAPLSSCSELQVSQFIRKHAVASCLLVSDTSTELSYILPSEAAKGA FERL FQHLE R	1298
Db	1170	DHESDTLTIDVS--AISNLIRKHVSEARLVEDIGHELTYVLPYEAKEGAFVELFHEIDD	1227
Qy	1299	SLDALHLSSFGIMDTTLEEVLKVSEEDQSLENSEADV KESRKDVLPGAEGPASGEGHAG	1358
Db	1228	RLSDLGISSYGISETTLEEIFLKVAEE-----SGVDA-ETSDGTLP-----	1267
Qy	1359	NLARCSELTQSQASLQSASSVGSARGDEGAGYTDVYGDYRPLF-DNPQDPD--NVSLQEV	1415
Db	1268	-----ARRNRRRA-FGDKQSCLRPFTEDDAADPNDSIDDPESR	1303
Qy	1416	EAEALSRV-GQGSRKLDGGWLKVRQFHGLLVKRFHCARRNSKALFSQIILLPAFFV CVAMT	1474
Db	1304	ETDLLSGMDGKGSYQVKGWKLTQQQFVALLWKRLLIARRSRKGFFAQIVLPAVFVCI ALV	1363
Qy	1475	VALSVPEIGDLPLPLVLSPSQYH-NYTQPRGNFIPYANEERREYRLRLSPDASPQQLVSTF	1533
Db	1364	FSLIVPPFGKYPSELELQPPWYNEQYT-----FVSNDAP E-----DTGTLELLNAL	1408
Qy	1534	RLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPP	1593
Db	1409	TKDPGFGTRCM-----EGNPI-----	1424
Qy	1594	PSPAPSDSPASPD EDLQAWNVS L PPTAGPEMWT SAPSLPRLVREPVR-----C	1641
Db	1425	-----PD-----TPCQAGEE EWTTAP-VPQTIMDLFQNGNWTM QNPSPAC	1463
Qy	1642	TCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILTDITGHN VSEYLLFTSDRF-----	1690
Db	1464	QCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYL VKTYVQIIAKSLKN	1523
Qy	1691	----RLHRYGAITFG--NVLKSIPASFGRAPPMVRK-----	1721
Db	1524	KIWNNEFRYGGFSLGVSNTQALPPSQEVND AIKQMKKHLKLAKDSSADRFNL SLGRFMTG	1583
Qy	1722	IAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSAS	1781
Db	1584	LDTKNNVKVWFNNKGWHAISSFLNVINNAILRANLQKGE-NPSHYGITAFNHPLNLTKQQ	1642
Qy	1782	LS-LDYLLQGTDVVIAIFIIIVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLAN	1840
Db	1643	LSEVALMTTSVDVLVSICVIFAMSFVPASFVFLIQERVSKAKHLQFISGVKPVIIYWLSN	1702
Qy	1841	YVWDMNLNYLVPATCCVIIIFVFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFVFEV	1900
Db	1703	FVWDMCNYVVEATLVIIIFICFQQKSYVSSTNLPLVLALLLLLYGWSITPLMYPASFVFKI	1762
Qy	1901	PSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLME	1960

```

Db      1763 PSTAYVVLTSVNLFINGSVATFVLELFT-DNKLNNINDILKSVFLIFPHFCLGRGLID 1821
Qy      1961 MAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQR 2020
      | |: : : : |: : : || ||:| | | ||||| ||:|: || | |:
Db      1822 MVKNQAMADALERFGE-NRFVSPLSWDLVGRNLFAMAVEGVVFFLITVLIQYRFFIRPRP 1880
Qy      2021 MPVSTKPVED-DVDVASERQVRVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVR 2079
      : | : | | | |||:| | ||:|:| ||:|: : | |||:|:|:
Db      1881 VNAKLSPLNDEDEDVRRERQRILDGGGQNDILEIKELTKIYRRK---RKPAVDRICVGIP 1937
Qy      2080 PGEFCGLLGVNGAGKTSTFKMLTGDESTTGGEAFVNGHSLVKELLQVQQSLGYCPQCDAL 2139
      ||||| |||||:||||| | : | |:|:| |:| : :| |:| ||| |:
Db      1938 PGEFCGLLGVNGAGKSSTFKMLTGDTTVTRGDAFLNKNLSILSNIHEVHQNMGYCPQFDAI 1997
Qy      2140 FDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAI 2199
      : || |||: : |||: |: |:|: || | ||:| || ||||| |:
Db      1998 TELLTGREHVEFFALLRGVPEKEVGKVGWEAIRKGLVKYGEKYAGNYSGGNKRKLSTAM 2057
Qy      2200 ALIGYPAFIFLDEPTTGMDPKARRFLWNLIIDLITGRSVVLTSHSMEECEALCTRLAIM 2259
      |||| | :||||| ||||| | :| ||||| |||||:|
Db      2058 ALIGGPPVFLDEPTTGMDPKARRFLWNCALSVVKEGRSVVLTSHSMEECEALCTRMAM 2117
Qy      2260 VNRLRLCLGSIQHLKNRFGDGYMITVR-TKSSQSVKDVVRFFNRFPEAMLKERHHTKVQ 2318
      |||| ||||:||||| | || |: |:| || ||:|:| |:
Db      2118 VNGRFRCCLGSVQHLKNRFGDGYTIVVRIAGSNPDLKPVQDFFGLAFPGSVLKEKHRNMLQ 2177
Qy      2319 YQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFVNFAKKQSDN 2368
      ||| | |||:|: | | | ||||| ||||| |||:
Db      2178 YQLPSSLSSLARIFSILSQSKRLHIEDYSVSQTTLDQVFNFAKDQSD 2227

```

RESULT 15

US-10-313-641-10

; Sequence 10, Application US/10313641

; Publication No. US20030162758A1

; GENERAL INFORMATION:

; APPLICANT: Ishida, Brian

; APPLICANT: Duncan, Keith

; APPLICANT: Bailey, Kathy

; APPLICANT: Kane, John

; APPLICANT: Schwartz, Daniel

; TITLE OF INVENTION: Treatments for Age Related-Macular Degeneration (AMD)

; FILE REFERENCE: P02351US2

; CURRENT APPLICATION NUMBER: US/10/313,641

; CURRENT FILING DATE: 2002-12-06

; PRIOR APPLICATION NUMBER: US 60/415,864

; PRIOR FILING DATE: 2002-10-03

; PRIOR APPLICATION NUMBER: US 60/340,498

; PRIOR FILING DATE: 2001-12-07

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 10

; LENGTH: 2261

; TYPE: PRT

; ORGANISM: Human

US-10-313-641-10

Query Match 33.5%; Score 4237.5; DB 14; Length 2261;  
 Best Local Similarity 39.8%; Pred. No. 0;  
 Matches 999; Conservative 346; Mismatches 730; Indels 435; Gaps 61;

Qy	6	QLQLLLWKNVTLKRRSPWVLAFEIFIPLVLFFILLGLRQKKPTISVKEVPFYTAAPL TSA	65
		:           :   :       :     :     :     :	
Db	6	QLRLLLWKNLTFRRRQTCQLLLEVAWPLFIFILILISVRLSYPPYEQHECHFPNKA-MPSA	64
Qy	66	GILPVMQSLCPDQGRDEFGFL-----QYANSTVTQLLERLDRVVEEGNLFDPARP	115
		:   : : :   : :     :     :   :   :	
Db	65	GTLPWVQGIICNANNPCFRYPTPGEAPGVVGNFNKSIVARLFSDARRLL----LYSQKDT	120
Qy	116	SLGSELEALR--QHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNL SLP	173
		: :       : :   : :       :	
Db	121	SMKDMRKVLR TLQ QIKKSSSNLKLQDFLVDNETFSG-----FLYHNLSLP	165
Qy	174	NSTAQALLAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELL LAPA	233
		:       :   :           :     :	
Db	166	KSTVDKMLRADV----ILHKVFLQGYQLHLTS-LCNGS-----KSEEMI----	204
Qy	234	LLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAELRNQLDV	293
		: : :   :   :   :     :	
Db	205	---QL---GDQEVSEL CGLPREKLAAAE-----RVLR SNMDI	235
Qy	294	AK-VSQQLGLDAPNGSDSSPQAPPPRRLQALLGDLLD-----AQKVLQDQDVLS	341
		: :   :     :   :             :   :   :	
Db	236	LKPILRTLNSTSPFPSKELAEA--TKTLLHSLGTLAQELFSMRSWSDMRQEV MFLT NVNS	293
Qy	342	ALALLLPQGACTGRTPGPPASGAGGAAN-----GTGAGAVMGP NATAE EGAPSAAALATP	396
		: :   :       : :   :       : :	
Db	294	SSSSTQIYQAVSRIVCGHPEGGLKIKSLNWIYEDNNYKALFGGNGTEEDAETFYDNSTTP	353
Qy	397	---DTLQGQCSAFVQ--LWAGLQPILCGNRTIEPEALRRGNMSSLGFTSKEQRNLGLIV	451
		: :   : : :     :   :	
Db	354	YCNDLMKNLESSPLSRIIWKALKPLLVG-----	381
Qy	452	HLMTSNPKILYAPAGSEVDRVILKANETFAFVG NVTHYAQVWLNISAEIRSFLEQGR LQQ	511
		: : :   :   : :   :   :   :	
Db	382	-----KILYTPDTPATRQVMAEVNKTFOELAVFHDLEG MWEE LSPKIWTFMENSQEMD	434
Qy	512	HLRWL-----QQYVAELRLHPE--ALNLSLDELPPALRQDNFS	547
		:                   :   :   :   :	
Db	435	LVRMLLDSRDNDHFWEQQLDGLDWTQAQDIVAFLAKHPEDVQSSNGSVYTWREAFNETN--	492
Qy	548	LPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQDNVTVFAS	607
		:     :       : : : :   : :   :   :	
Db	493	-----QAIRTIS-----RFMECVNLNKLEPIATEVWLINKSME--LLDERKFWAG	535
Qy	608	VIFQTRKDGS--LPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTG---GRFYFLYGFVW	662
		: :             : :   :   :           :	
Db	536	IVFTGITPGSIELPHHVKYKIRMDIDNVERTNKIKDGYWDPGPRADPFEDMRYVWGGFAY	595
Qy	663	IQDMMERAIIDTFVGHDVVEPGSYVQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSVAM	722
		:   : :   :     : :   :               :     :   :   :	
Db	596	LQDVVEQAIIRVLTGTE-KKTGVYMQMPYPCYVDDIFLRVMSRSMPLFMTLAWIYSVAV	654

Qy	723	TIQHIVAEKEHRLKEVMKTMGLNNAVHWVAFITGFVQLSISVTALTAILKYGQVLMHSH	782
Db	655	IIKGIVYEKEARLKETMRIMGLDNSILWFSWFISSLIPLLVSAAGLLVVLKLGNNLLPYSD	714
Qy	783	VVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHDK	842
Db	715	PSVVFVFLSVFAVVTILQCFLISTLFSRANLAAACGGIIYFTLYLPYVLC-----VAWQD	769
Qy	843	ITAFE-KCIASLMSTTAFGLGSKYFALYEAVGVGIQWHTFSQSPVEGDDFNLLLAVTMLM	901
Db	770	YVGFTLKIFASLLSPVAFGFGCEYFALFEEQGIGVQWDNLFESPVEEDGFNLTTSVSMML	829
Qy	902	VDAVVGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPPWARTPRLSVM	961
Db	830	FDTFLYGVMTWYIEAVFPGQYGI PRPWYFPCTKSYWFG-----ESDEKSHPGSNQKRIS--	884
Qy	962	EEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTKVYKDDKKLALNKLNLNYENQVV	1021
Db	885	---EIC-----MEEEPHTLKLGVSIQNLVKVYRDGMKVAVDGLALNFYEGQIT	929
Qy	1022	SFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLT	1081
Db	930	SFLGHNGAGKTTTMSILTGLFPPTSGTAYILGKDIRSEMSTIRQNLGVCPPQHNVLFDMILT	989
Qy	1082	VEEHLWFYSRLKSMAQEEIRREMDKMIEDLEL-SNKRHSLVQTLSGGMKRKLSVAIAFVG	1140
Db	990	VEEHIWFYARLKLSEKHVKAEMEQMALDVGLPSSSKLKSQTSQLSGGMQRKLSVALAFVG	1049
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHMHDEADLLGDRIAIISHGKL	1200
Db	1050	GSKVVILDEPTAGVDPYSRRGIWELLKYRQGRITILLSTHMHDEADVLGDRIAIISHGKL	1109
Qy	1201	KCCGSPLFLKGTYG DGYRLTLVKRPAEPG-----GPQEPGLAS	1238
Db	1110	CCVGSSFLKNQLGTGYRLTLVKKDVESSLSSCRNSSSTVSYLKKEDEVSSQSSSDAGLGS	1169
Qy	1239	SPPGRAPLSSCSELQVSQFIRKHHVASCLLVSDTSTELSYILPSEAAKGAERLFOHLER	1298
Db	1170	DHESDTLTIDVS--AISNLIRKHVSEARLVEDIGHELTYVLPYEAKEGAFVELFHEIDD	1227
Qy	1299	SLDALHLSSFGLMDTTLEEVLKVSEEDQSLENSEADVKE SRKDVLPGAEGPASGEGHAG	1358
Db	1228	RLSDLGISSYGISETTLEEIFLKVAEE-----SGVDA-ETSDGTLP-----	1267
Qy	1359	NLARCSELTQSQASLQSASSVGSARGDEGAGYTDVYGDYRPLF-DNPQDPD--NVSLQEV	1415
Db	1268	-----ARRNRRA-FGDKQSCLRPFTEDDAADPNDSIDIPESR	1303
Qy	1416	EAEALS RV-GQGSRKLDGGWLKVRQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMT	1474
Db	1304	ETDLLSGMDGKGSYQVKGWKLTQQQFVALLWKRLLIARRSRKGFFAQIVLPAVFCIALV	1363
Qy	1475	VALSVPEIGDLPLVLSPSQYH-NYTQPRGNFIPYANEERREYRLRLSPDASPQQLVSTF	1533
Db	1364	FSLIVPPFGKYPSELQLPWWYNEQYT-----FVSNDAP-----DTGTLELLNAL	1408
Qy	1534	RLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFDSMCLESFTOGLPLSNFVPPP	1593

Db	1409	TKDPGFGTRCM-----EGNFI-----	1424
Qy	1594	PSPAPSDSPASPEDLQAWNVS L PPTAGPEMWT SAPSLPRLVREPVR-----C	1641
Db	1425	-----PD-----TPCQAGEEWT TAP-VPQTIMDLFQNGNWTMQNPSPAC	1463
Qy	1642	TCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILTDITGHNVS EYLLFTSDRF-----	1690
Db	1464	QCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKN	1523
Qy	1691	----RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK-----	1721
Db	1524	KIWNNEFRYGGFSLGVSNTQALPPSQEVND AIKQMKHLKLAKDSSADRFLNSLGRFMTG	1583
Qy	1722	IAVRRAAQVFYNNKGYHSMPTYLNSLNAILRANLPKSKGNPAAYGITVTNHPMNKTSAS	1781
Db	1584	LDTKNNVKVWFNNKGWHAISSFLNVINAILRANLQKGE-NPSHYGITAFNHPLNLTKQQ	1642
Qy	1782	LS-LDYLLQGTDVVIAIFIIVAMS FVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLAN	1840
Db	1643	LSEVALMTTSVDVLVSICVIFAMSFVPASFVFLIQERVSKAKHLQFISGVKPVIIYWSN	1702
Qy	1841	YVWDMNLNYLVPATCCVIIIFVFDLPAYTSPTNFPAVLSLFLLYGWSITPIMYPASFWFEV	1900
Db	1703	FVWDMCNVVPATLVIIIFICFQQKSYSSTNLPVLALLLLLYGWSITPLMYPASFVFKI	1762
Qy	1901	PSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLME	1960
Db	1763	PSTAYVVLTSVNLFIGINGSVATFVLELEFT-DNKLNNINDILKSVFLIFPHFCLGRGLID	1821
Qy	1961	MAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQR	2020
Db	1822	MVKNQAMADALERFGE-NRFVSPLSWDLVGRNLFAMAVEGVVFFLITVLIQYRFFIRPRP	1880
Qy	2021	MPVSTKPVED-DVDVASERQVR LGDADNDMVKIENLTKVYKSRKIGRILAVDRCLCGVR	2079
Db	1881	VNAKLSPLNDEDEDVRRERQRILDGGGQNDILEIKELTKIYRRK---RKPAVDRICVGIP	1937
Qy	2080	PGECFGLLG VNGAGKTSTFKMLTGDESTTGGEAFVNGH SVLKELLQVQQSLGYCPQCDAL	2139
Db	1938	PGECFGLLG VNGAGKSSTFKMLTGDTTVTRGDAFLNKNSILSNIHEVHQNMGYCPQFDAI	1997
Qy	2140	FDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELT KYADKPAGTYSGGNKRKLSTAI	2199
Db	1998	TELLTGREHVEFFALLRGVPEKEVGKVG EWAIRKGLVKYGEKYAGNYSGGNKRKLSTAM	2057
Qy	2200	ALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIM	2259
Db	2058	ALIGGPPVFLDEPTTGMDPKARRFLWNCALSVVKEGRSVVLTSHSMEECEALCTRM AIM	2117
Qy	2260	VNGRLRCLGSIQHLKNRFGDGYMITVR-TKSSQSVKDVVRFFNRNFPEAM LKERHHTKVQ	2318
Db	2118	VNGRFRCLG SVQH LKNRFGDGYTIVVRIAGSNPDLKPVDFFGLAFPGSVLKEKH RNMQLQ	2177
Qy	2319	YQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFNFAKKQSDN	2368



Db 2178 YQLPSSLSSLARIFSILSQSKKRLHIEDYSVSQTLLDQVFVNFAKDQSDD 2227

Search completed: September 1, 2004, 11:12:58  
Job time : 234 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2004, 10:44:52 ; Search time 209 Seconds  
(without alignments)  
3677.523 Million cell updates/sec

Title: US-10-088-467-2  
Perfect score: 12668  
Sequence: 1 MGFLHQLQLLLWKNVTLKRR.....GLISFEEERAQLSFNTDTLC 2436

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query					
No.	Score	Match	Length	DB	ID	Description
-----						

1	12668	100.0	2436	4	Q9HC28	Q9hc28 homo sapien
2	11725	92.6	2434	11	Q9ESR9	Q9esr9 rattus norv
3	4518	35.7	867	4	Q96HC2	Q96hc2 homo sapien
4	4214	33.3	2260	13	Q8UVV4	Q8uvv4 gallus gall
5	4083	32.2	2201	11	Q80ZB2	Q80zb2 rattus norv
6	3951	31.2	2310	11	O35600	O35600 mus musculu
7	3893.5	30.7	2281	6	O02698	O02698 bos taurus
8	3681	29.1	2170	11	Q7TNJ2	Q7tnj2 rattus norv
9	3675	29.0	2159	11	Q91V24	Q91v24 mus musculu
10	3620	28.6	2146	4	Q9BZC4	Q9bzc4 homo sapien
11	3618	28.6	2146	4	Q8IZY2	Q8izy2 homo sapien
12	3616	28.5	2146	4	Q9NR73	Q9nr73 homo sapien
13	3515	27.7	2008	4	Q96S58	Q96s58 homo sapien
14	2831.5	22.4	2595	4	Q86UK0	Q86uk0 homo sapien
15	2827.5	22.3	2347	4	Q8IZW6	Q8izw6 homo sapien
16	2821.5	22.3	2277	4	Q96JT3	Q96jt3 homo sapien
17	2687.5	21.2	5058	4	Q86UQ4	Q86uq4 homo sapien
18	2679	21.1	2127	4	Q86WI2	Q86wi2 homo sapien
19	2645.5	20.9	2143	11	Q80T20	Q80t20 mus musculu
20	2622.5	20.7	1764	5	Q8MUA3	Q8mua3 strongyloce
21	2582.5	20.4	1704	11	Q8R420	Q8r420 mus musculu
22	2561	20.2	1538	11	Q80XT2	Q80xt2 mus musculu
23	2313.5	18.3	1713	5	Q9VRG4	Q9vrg4 drosophila
24	2262.5	17.9	1487	11	Q8BPY1	Q8bpy1 mus musculu
25	2127.5	16.8	1882	10	Q8W010	Q8w010 arabidopsis
26	2024	16.0	1802	5	Q9TXV8	Q9txv8 caenorhabdi
27	2009	15.9	1660	5	Q9VVJ9	Q9vvj9 drosophila
28	1971	15.6	1896	5	Q8WS98	Q8ws98 leishmania
29	1964.5	15.5	1816	10	Q84M24	Q84m24 arabidopsis
30	1920	15.2	373	4	Q9UPU0	Q9upu0 homo sapien
31	1897.5	15.0	1547	5	O01790	O01790 caenorhabdi
32	1882	14.9	1843	5	Q9GQS2	Q9gqs2 leishmania
33	1762	13.9	1621	5	Q8T6J5	Q8t6j5 dictyosteli
34	1732.5	13.7	1702	5	Q8T6J4	Q8t6j4 dictyosteli
35	1710.5	13.5	1631	5	Q8T6J1	Q8t6j1 dictyosteli
36	1709.5	13.5	1750	5	Q9BKL1	Q9bkl1 trypanosoma
37	1694	13.4	1608	5	Q8T6J3	Q8t6j3 dictyosteli
38	1648	13.0	1591	5	Q86IX7	Q86ix7 dictyosteli
39	1645.5	13.0	1662	5	Q8T6J2	Q8t6j2 dictyosteli
40	1473.5	11.6	1222	10	Q8S8T6	Q8s8t6 arabidopsis
41	1468.5	11.6	1620	11	Q8K440	Q8k440 mus musculu
42	1461	11.5	1642	4	Q9NY14	Q9nyl4 homo sapien
43	1459	11.5	1642	4	Q8WWZ7	Q8wwz7 homo sapien
44	1457.5	11.5	1642	11	Q8CF82	Q8cf82 rattus norv
45	1448.5	11.4	1623	11	Q8K449	Q8k449 mus musculu

# ALIGNMENTS

## RESULT 1

Q9HC28

ID Q9HC28 PRELIMINARY; PRT; 2436 AA.  
AC Q9HC28;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE ATP-binding cassette sub-family A member 2 (ABC transporter  
 DE ABCA2).  
 GN ABCA2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Vulevic B., Chen Z., Davis W. Jr., Walsh E.S., Tew K.D.;  
 RT "Cloning and characterization of human ABCA2.";  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=11178988;  
 RA Kaminski W.E., Piehler A., Pullmann K., Porsch-Ozcurumez M., Duong C.,  
 RA Bared G.M., Buchler C., Schmitz G.;  
 RT "Complete Coding Sequence, Promoter Region, and Genomic Structure of  
 RT the Human ABCA2 Gene and Evidence for Sterol-Dependent Regulation in  
 RT Macrophages.";  
 RL Biochem. Biophys. Res. Commun. 281:249-258(2001).  
 DR EMBL; AF178941; AAG09372.1; -.  
 DR EMBL; AF327657; AAK14334.1; -.  
 DR PIR; A59189; A59189.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR000566; Lipocln\_cytFABP.  
 DR InterPro; IPR000572; Oxidored\_molyb.  
 DR Pfam; PF00005; ABC\_tran; 2.  
 DR ProDom; PD000006; ABC\_transporter; 2.  
 DR SMART; SM00382; AAA; 2.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS00213; LIPOCALIN; 1.  
 DR PROSITE; PS00559; MOLYBDOPTERIN\_EUK; 1.  
 KW ATP-binding.  
 SQ SEQUENCE 2436 AA; 269955 MW; E044A3AF14EA25D1 CRC64;

Query Match 100.0%; Score 12668; DB 4; Length 2436;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGFLHQLQLLLWKNVTLKRRSPWVLA FEIFIP LVLF FILLGLRQKKPTISVKEVPFYTAA 60  
 |||  
 Db 1 MGFLHQLQLLLWKNVTLKRRSPWVLA FEIFIP LVLF FILLGLRQKKPTISVKEVPFYTAA 60  
 Qy 61 PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLDRVVEEGNLFDPARPSLGSE 120

Db	61	 PLTSAGILPVMQSLCPDQGRDEFGLQYANSTVTQLLERLDRVVEEGLNLFDPARPSLGSE	120
Qy	121	LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSPNSTAQAL	180
Db	121	LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSPNSTAQAL	180
Qy	181	LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPALLEQLTC	240
Db	181	LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPALLEQLTC	240
Qy	241	TPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAELRNQLDVAKVSQQ	300
Db	241	TPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAELRNQLDVAKVSQQ	300
Qy	301	GLDAPNGSDSSPQAPPPRRLQALLGDLLDAQVQLQDVVLSALALLPQGACTGRTPGPP	360
Db	301	GLDAPNGSDSSPQAPPPRRLQALLGDLLDAQVQLQDVVLSALALLPQGACTGRTPGPP	360
Qy	361	ASGAGGAANGTGAGAVMGPNATAEEGAPSAAALATPDTLQGQCSAFVQLWAGLQPILCGN	420
Db	361	ASGAGGAANGTGAGAVMGPNATAEEGAPSAAALATPDTLQGQCSAFVQLWAGLQPILCGN	420
Qy	421	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF	480
Db	421	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF	480
Qy	481	AFVGNVTHYAQVWLNISAEIRSFLQGRLLQQLRWLQYVAELRLHPEALNLSDELPPA	540
Db	481	AFVGNVTHYAQVWLNISAEIRSFLQGRLLQQLRWLQYVAELRLHPEALNLSDELPPA	540
Qy	541	LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPEESIVNYTLNQAYQD	600
Db	541	LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPEESIVNYTLNQAYQD	600
Qy	601	NVTVFASVIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGREFYFLYGF	660
Db	601	NVTVFASVIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGREFYFLYGF	660
Qy	661	VWIQDMMERAIIDTFVGHVVEPGSYVQMFYPYCYTRDDFLFVIEHMMPLCMVISWVYSV	720
Db	661	VWIQDMMERAIIDTFVGHVVEPGSYVQMFYPYCYTRDDFLFVIEHMMPLCMVISWVYSV	720
Qy	721	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWWAWFITGFVQLSISVTALTALILKYGQVLMH	780
Db	721	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWWAWFITGFVQLSISVTALTALILKYGQVLMH	780
Qy	781	SHVVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAH	840
Db	781	SHVVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAH	840
Qy	841	DKITAFEKCIASLMSTTAFGLGSKYFALYEAVGVGIQWHTFSQSPVEGDDFNLLAVTML	900
Db	841	DKITAFEKCIASLMSTTAFGLGSKYFALYEAVGVGIQWHTFSQSPVEGDDFNLLAVTML	900
Qy	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSV	960

Db	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSV	960
Qy	961	MEEDQACAMESRRFEETRGMEEETHLPLVVCVDKLTQVYKDDKKLALNKLNLNLYENQV	1020
Db	961	MEEDQACAMESRRFEETRGMEEETHLPLVVCVDKLTQVYKDDKKLALNKLNLNLYENQV	1020
Qy	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR	1080
Db	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR	1080
Qy	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAFVG	1140
Db	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAFVG	1140
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Db	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Qy	1201	KCCGSPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Db	1201	KCCGSPLFLKGTYGDGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Qy	1261	HVASCLLVSDTSTELSYILPSEAAKKGAFERLFQHLERSLDALHLSSFGLMDTTLEEVFL	1320
Db	1261	HVASCLLVSDTSTELSYILPSEAAKKGAFERLFQHLERSLDALHLSSFGLMDTTLEEVFL	1320
Qy	1321	KVSEEDQSLENSEADVKEARKDVLPGAEGPASGEGHAGNLARCSELTQSQASLQSASSVG	1380
Db	1321	KVSEEDQSLENSEADVKEARKDVLPGAEGPASGEGHAGNLARCSELTQSQASLQSASSVG	1380
Qy	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALS RVGQGSRKLDGGWLKVRQF	1440
Db	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALS RVGQGSRKLDGGWLKVRQF	1440
Qy	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSYHNYTQ	1500
Db	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSYHNYTQ	1500
Qy	1501	PRGNFIPIYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Db	1501	PRGNFIPIYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Qy	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS LPPTA	1620
Db	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVS LPPTA	1620
Qy	1621	GPEMWTSAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTITGHNV	1680
Db	1621	GPEMWTSAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTITGHNV	1680
Qy	1681	EYLLFTSDRFLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYSM	1740
Db	1681	EYLLFTSDRFLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYSM	1740
Qy	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFII	1800
Db	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFII	1800

Qy	1801	VAMSFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVIILF	1860
Db	1801	VAMSFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVIILF	1860
Qy	1861	VFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGITAT	1920
Db	1861	VFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGITAT	1920
Qy	1921	VATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDM	1980
Db	1921	VATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDM	1980
Qy	1981	KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDQDVASERQR	2040
Db	1981	KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDQDVASERQR	2040
Qy	2041	VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFLLGVNGAGKTSTFKM	2100
Db	2041	VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFLLGVNGAGKTSTFKM	2100
Qy	2101	LTGDESTTGGEAFVNGHSLKELLQVQOSLGYCPQCDALFDELTAREHLQLYTRLRGISW	2160
Db	2101	LTGDESTTGGEAFVNGHSLKELLQVQOSLGYCPQCDALFDELTAREHLQLYTRLRGISW	2160
Qy	2161	KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK	2220
Db	2161	KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK	2220
Qy	2221	ARRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG	2280
Db	2221	ARRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG	2280
Qy	2281	YMITVRTKSSQSVKDVVRFFNRFPEAMLERHHTKVQYQLKSEHISLAQVFSKMEQVSG	2340
Db	2281	YMITVRTKSSQSVKDVVRFFNRFPEAMLERHHTKVQYQLKSEHISLAQVFSKMEQVSG	2340
Qy	2341	VLGIEDYSVSQTTLDNVFVNFAKKQSDNLEQQETEPSPALQSPLGCLLSLLRPRSAPTEL	2400
Db	2341	VLGIEDYSVSQTTLDNVFVNFAKKQSDNLEQQETEPSPALQSPLGCLLSLLRPRSAPTEL	2400
Qy	2401	RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC	2436
Db	2401	RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC	2436

# RESULT 2

Q9ESR9

ID Q9ESR9 PRELIMINARY; PRT; 2434 AA.

AC Q9ESR9;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE ABC2.

GN ABC2.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=20427713; PubMed=10970803;  
 RA Zhao L., Zhou C., Tanaka A., Nakata M., Hirabayashi T., Amachi T.,  
 RA Shioda S., Ueda K., Inagaki N.;  
 RT "Cloning, characterization and tissue distribution of the rat ATP-  
 RT binding cassette (ABC) transporter ABC2/ABCA2.";  
 RL Biochem. J. 350:865-872(2000).  
 DR EMBL; AB037937; BAB16596.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR000566; Lipocln\_cytFABP.  
 DR Pfam; PF00005; ABC\_tran; 2.  
 DR ProDom; PD000006; ABC\_transporter; 2.  
 DR SMART; SM00382; AAA; 2.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS00213; LIPOCALIN; 1.  
 KW ATP-binding.  
 SQ SEQUENCE 2434 AA; 270925 MW; CD424A9C4F63513F CRC64;

Query Match 92.6%; Score 11725; DB 11; Length 2434;  
 Best Local Similarity 92.8%; Pred. No. 0;  
 Matches 2262; Conservative 49; Mismatches 122; Indels 4; Gaps 4;

Qy	1	MGFLHQQLQLLLWKNVTLKRRSPWVLAFEIFIPVLVFFILLGLRQKKPTISVKEVPFYTAA	60
Db	1	MGFLHQQLQLLLWKNVTLKRRSPWVLAFEIFIPVLVFFILLGLRQKKPTISVKEA-FYTAA	59
Qy	61	PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLDRVVEEGLNLFDPARPSLGSE	120
Db	60	PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLNRVVEESNLFDPERPSLGSE	119
Qy	121	LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSPNSTAQAL	180
Db	120	LEALHQRLLEALSSGPGTWESHARPVSSFSLDSVARDKRELWRFLMQNLSPNSTAQAL	179
Qy	181	LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLAPALLEQLTC	240
Db	180	LAARVDPSEVYRLLFGPLPDLDGKLGFLRKQEPWSHLGSNPLFQMEELLAPALLEQLTC	239
Qy	241	TPGSSELGRILTVPEQKQALQGYRDAVCSGQAAARARRFSGLSAELRNQLDVAKVSQQL	300
Db	240	APGSSELGRILTMPEGHQVDLQGYRDAVCSGQATARAQHFSDIATELRNQLDIAKIAQQL	299
Qy	301	GLDAPNGSDSSPQAPPRLQALLGDLLDAQKVLQDQVDVLSALALLLPQGACTGRTPGPP	360



[illegible]

Db	1140	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1199
Qy	1201	KCCGSPLFLKGTGYGDYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Db	1200	:          :    :	
Db	1200	KCCGSPLFLKGAYGDYRLTLVKRPAEPGTSQEPGMASSPSGRPQLSNCSEMQVSQFIRK	1259
Qy	1261	HVASCLLVSDTSTELSYILPSEAAKKGAFERLFQHLERSLDALHLSSFGMLDTTLEEVFL	1320
Db	1260		
Db	1260	HVASSLLVSDTSTELSYILPSEAVKKGAERLFQQLEHSLDALHLSSFGMLDTTLEEVFL	1319
Qy	1321	KVSEEDQSLENSEADVKE SRKDVLPGAEGPASGEGHAGNLARCELTQSQASLQSASSVG	1380
Db	1320	:	
Db	1320	KVSEEDQSLENSEADVKE SRKDALPGAEGLTAVESQAGNLARCELAQSQASLQSASSVG	1379
Qy	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVQGSRKLDGGWLKVRQF	1440
Db	1380	:          :       :    :	
Db	1380	SARGDEGAGYTDGYGDYRPLFDNLQDPDSVSLQEAEALARVQGSRKLEGWWLKMRQF	1439
Qy	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQ	1500
Db	1440		
Db	1440	HGLLVKRFHCARRNSKALCSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYHNYTQ	1499
Qy	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Db	1500		
Db	1500	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPMNLNS	1559
Qy	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDED-LQAWNVS LPPT	1619
Db	1560		
Db	1560	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPLSPDEDSLLAWNTSLPPT	1619
Qy	1620	AGPEMWTSA PSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTDTIGHNV	1679
Db	1620		
Db	1620	AGPETWTWAPSLPRLVHEPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTDTIGHNV	1679
Qy	1680	SEYLLFTSDRFLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYS	1739
Db	1680	:	
Db	1680	SEYLLFTSDRFLHRYGAITFGNIQKSIPAPIGTRTPLMVRKIAVRRVAQVLYNNKGYS	1739
Qy	1740	MPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFI	1799
Db	1740		
Db	1740	MPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFI	1799
Qy	1800	IVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVIIIL	1859
Db	1800		
Db	1800	IVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCIIIL	1859
Qy	1860	FVFDLPAYTSPTNFPVAVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITA	1919
Db	1860		
Db	1860	FVFDLPAYTSPTNFPVAVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITA	1919
Qy	1920	TVATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDK	1979
Db	1920		
Db	1920	TVATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEIAYNEYINEYYAKIGQFDK	1979
Qy	1980	MKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDQDVASERQ	2039
Db	1980		
Db	1980	MKSPFEWDIVTRGLVAMTVEGFVGVGFLLTIMCQYNFLRQPQRLPVSTKPVEDDQDVASERQ	2039

Qy 2040 RVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLGVNGAGKTSTFK 2099  
 |||  
 Db 2040 RVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLGVNGAGKTSTFK 2099  
 |||  
 Qy 2100 MLTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGIS 2159  
 |||:|  
 Db 2100 MLTGDESTTGGEAFVNGHSVLKDLLQVQQSLGYCPQFDALFDELTAREHLQLYTRLRGIP 2159  
 |||:|  
 Qy 2160 WKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDP 2219  
 |||:|  
 Db 2160 WKDEAQVVRWALEKLELTKCADKPAGSYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDP 2219  
 |||:|  
 Qy 2220 KARRFLWNLILDLIKTRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGD 2279  
 |||:|  
 Db 2220 KARRFLWNLILDLIKTRSVVLTSHSMEECEAVCTRLAIMVNGRLRCLGSIQHLKNRFGD 2279  
 |||:|  
 Qy 2280 GYMITVRTKSSQSVKDVVRFFNRNFP EAM LKERHHTKVQYQLKSEHISLAQVFSKMEQVS 2339  
 |||:|  
 Db 2280 GYMITVRTKSSQNVKDVVRFFNRNFP EAM LKERHHTKVQYQLKSEHISLAQVFSKMEHV 2339  
 |||:|  
 Qy 2340 GVLGIEDYSVSQTTLDNVFVNFAKKQSDNLEQQETEPPSALQSPLGCLLSLLRPRAPTE 2399  
 |||:|  
 Db 2340 GVLGIEDYSVSQTTLDNVFVNFAKKQSDNVEQQEAE-PSTLPSPLG-LLSLLRPRAPTE 2397  
 |||:|  
 Qy 2400 LRALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2436  
 |||  
 Db 2398 LRALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2434  
 |||

# RESULT 3

Q96HC2

ID Q96HC2 PRELIMINARY; PRT; 867 AA.  
 AC Q96HC2;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Similar to KIAA1062 protein (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Eye;  
 RA Strausberg R.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC008755; AAH08755.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR006209; EGF\_like.  
 DR Pfam; PF00005; ABC\_tran; 1.

DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 KW ATP-binding.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 867 AA; 96734 MW; DCF6B6A90074C085 CRC64;

Query Match 35.7%; Score 4518; DB 4; Length 867;  
 Best Local Similarity 99.9%; Pred. No. 3.6e-291;  
 Matches 866; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1570	RFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDELQAWNVS	LPPTAGPEMWTSAP	1629
Db	1	RFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDELQAWNVS	LPPTAGPEMWTSAP	60
Qy	1630	SLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTDITGHN	VSEYLLFTSDR	1689
Db	61	SLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVAGDILTDITGHN	VSEYLLFTSDR	120
Qy	1690	FRLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYH	SMPTYLNSLNN	1749
Db	121	FRLHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYH	SMPTYLNSLNN	180
Qy	1750	AILRANLPKSKGNPAAYGITVTNHMPNKTASLSLDYLLQGTDVVIAIFI	IIVAMSFVPAS	1809
Db	181	AILRANLPKSKGNPAAYGITVTNHMPNKTASLSLDYLLQGTDVVIAIFI	IIVAMSFVPAS	240
Qy	1810	FVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVII	ILFVFDLPAYTS	1869
Db	241	FVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVII	ILFVFDLPAYTS	300
Qy	1870	PTNFPAVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGIT	ATVATFLLQLF	1929
Db	301	PTNFPAVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGIT	ATVATFLLQLF	360
Qy	1930	EHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFD	KMKSPFEWDIV	1989
Db	361	EHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFD	KMKSPFEWDIV	420
Qy	1990	TRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDDVDVASE	RQVLRGDADND	2049
Db	421	TRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDDVDVASE	RQVLRGDADND	480
Qy	2050	MVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFLLGVNGAGKTSTF	KMLTGDESTTG	2109
Db	481	MVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFLLGVNGAGKTSTF	KMLTGDESTTG	540
Qy	2110	GEAFVNGHSLVKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRG	ISWKDEARVVKW	2169
Db	541	GEAFVNGHSLVKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRG	ISWKDEARVVKW	600
Qy	2170	ALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTG	MDPKARRFLWNLI	2229
Db	601	ALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTG	MDPKARRFLWNLI	660
Qy	2230	LDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNR	FGDGYMITVVRTKS	2289

```

      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 LDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVVRTKS 720
Qy      2290 SQSVKDVVRFFNRFPEAMKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSV 2349
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 SQSVKDVVRFFNRFPEAMKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSV 780
Qy      2350 SQTTLDNVFNFAKKQSDNLEQQETEPPSALQSPLGCLLSLLRPRSAPTELRLVADEPE 2409
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 SQTTLDNVFNFAKKQSDNLEQQETEPPSALQSPLGCLLSLLRPRSAPTELRLVADEPE 840
Qy      2410 DLDTEDEGLISFEEERAQLSFNTDTLC 2436
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 DLDTEDEGLISFEEERAQLSFNTDTLC 867

```

RESULT 4

Q8UVV4

```

ID   Q8UVV4      PRELIMINARY;      PRT;  2260 AA.
AC   Q8UVV4;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   ATP-binding cassette transporter 1.
GN   ABCA1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Attie A.D., Brooks-Wilson A.R., Walker D., McManus B.,
RA   Gray-Kellar M.P., MacDonald M.L.E., Roomp K., Tebon A., Zhang L.-H.,
RA   Mulligan J., Sensen C., Bitgood J.J., Cook M.E., Kastelein J.J.P.,
RA   Hayden M.R.;
RT   "Cholesterol Ester Accumulation in Hepatocytes and Intestinal Lamina
RT   Propria Caused by an ABCA1 Mutation in WHAM Chickens.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR   EMBL; AF362377; AAL56247.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transporter.
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transporter; 2.
DR   SMART; SM00382; AAA; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR   PROSITE; PS50893; ABC_TRANSPORTER_2; 2.
KW   ATP-binding.
SQ   SEQUENCE  2260 AA;  254070 MW;  19D137F342F98662 CRC64;

```

```

Query Match          33.3%;  Score 4214;  DB 13;  Length 2260;
Best Local Similarity 38.9%;  Pred. No. 3e-270;

```

Matches 982; Conservative 367; Mismatches 719; Indels 458; Gaps 60;

```
Qy      1 MGFLHQQLQLLWKNVTLKRRSPWVLAFEIFIPLVLFFILLGLRQKKPTISVKEVPFYTAA 60
      | | || ||||| | :|| : | |: || :|||: :| | | |
Db      1 MAFWTQLGLLLWKNFTYRRRQTFQLLIEVAWPLFIFFILISVRLSYPPYEQHECHFPA 60

Qy     61 PLTSAGILPVMQSLCPDGQRDEF-----GFLQYANSTVTQLLERLDRVVEEGNLF 111
      : ||| || :| : : | | : | :|: | |
Db     61 -MPSAGTLPWIQGIICNANNPCFRYPTPGESPGIVGNFNASIV-----SRLES 107

Qy    112 PARPSL-----GSELEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDVARNPQELWRFL 166
      |: | : : : : | | :|| | : | :|: : ||
Db    108 DAKRLLLYSQDTSIKDVQKVLAKLRKLGNSGLDL---KLRDFLVDN-----ETFSDFL 159

Qy    167 TQNLSLPNSTAQALLAARVDPPEV-----YHLLFGPSSALDSQS- 205
      |:|:|:| : || | |: :| : : | | :|:|:
Db    160 RHNVSMPSSAVEELLDAEVNLQKVIVSGYRIQLRDLNCSLSSEFLTQNRSVAMDSEAF 219

Qy    206 -----GLHKGQEPWSRLGGNPLFRMEELLLAPALLEQLTCTPGSGELGRILTVPESQK 258
      || : : | ||| : : : | |: |
Db    220 LCTLPKETLHAAELAF-RANLNPLKPLQREIFFNSSLRDLSET----- 261

Qy    259 GALQGYRDAVCSGQAAARARRFSGLSAELRNQLDVAKVSQQ-LGLDAPNGSDSSPQAPPP 317
      :: ||:: | || : : : |: : | | |:| |
Db    262 --VEALRDSL-----GKLVKELLSMKSWSDMRQEVMTNVDNASNSSTQI--- 304

Qy    318 RRLQALLGDLLDAQKVLQDQDVLSALALLLPQGACTGRTPGPPASGAGGAA-----NG 370
      :| : | :| | |
Db    305 -----YQAVSRIVCGHPEG-----GGLKIKSLNWDYED 331

Qy    371 TGAGAVMGNPNATAEEGAPSAAALATPDTLQGQCSAFVQ-----LWAGLQPILCGN 420
      |: | |:| :: || |: : : :| |:|:| |
Db    332 NNYKALFGGNSTEDDVTNFYDNSTTP-----YCNELMKNLESSPLSRIWRALKPLLIG- 385

Qy    421 NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHMTSNPKILYAPAGSEVDRVILKANETF 480
      | :|| | : : : : | ||
Db    386 -----KVLYTPDTPAIRKIMAEVNRTF 407

Qy    481 AFVGNVTHYAQVWLNISAEIRSFLEQGRLLQQLRWLQQYVAELRLHPEALNLSLDELPPA 540
      :| :| || :| :|:| : :| | : | || | :|:|
Db    408 QELGVFRDLGGMWEEISPKIWTFMESSQEMDLIRTLKSKALWDLHLPASNWTVEDVARF 467

Qy    541 LRQ--DNFSLPSGMAL--LQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQ 596
      | : : | :|| : : | | :|| |:| : | :| :|
Db    468 LSKHPEEFADNGMVYTWVDAFNETDRAIQTISRMECVNLDKLEPVATEVRLINKSLE- 526

Qy    597 AYQDNVTVFASVIFQTRKDG--LPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTG--- 651
      | :| |:| | || || ||| : |:|:|: || |||
Db    527 -LLDERRFWAGVVFTEIAPNSTELPQHVKYKIRMDIDNVERTNKKIDGYWDPGPRADPFE 585

Qy    652 GRFYFLYGFWIQDMMERAIIDTFVGHVVEPGSYVQMPYPYCYTRDDFLFVIEHMMPLC 711
      | || :|:|:|:| | : : | ||| |||| | || |: |||
Db    586 DMRYVWGGFTYLQDVVEQAIIRVQTGTE-KKTGVYVQQMPYPCYVDDIFLRVMSRSMPLF 644

Qy    712 MVISWVYSVAMTIQHIVA EKEHRLKEVMKTMGLNNAVHWVAFITGFVQLSISVTALTAI 771
      | :|:|:|:| :| || ||| ||| |: |||: : |:|:|: : | :| | |
Db    645 MTLAWIYSVAVIIKGIVYEKEARLKETMRIMGLDNGILWLSWFISSLIPLIMSAGLLVLI 704
```

Qy	772	LKYGQVLMHSHVVIWLFSLAVYAVATIMFCFLVSVLYSKAKLASACGGIIFYLSYVPYMY	831
Db	705	LKMGNLLPYSDPSVVFVFLSIFGIVTILQCFLISTVFSRANLAAACGGIVYFTLYLPYVL	764
Qy	832	VAIREEVAHDKITAFE-KCIASLMSTTAFGLGSKYFALYEYVAGVGIQWHTFSQSPVEGDD	890
Db	765	C-----VAWQDYVSFSLKIFASLLSPVAFGFGCEYFALFEEQGVGVQWDNFFESPLEDG	819
Qy	891	FNLLLAVTMLMVDVAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESW	950
Db	820	FSITTSAVMMLFDTFLYGVMTWYIESVFPGQYGI PRPWYF PFTKSYWFGE-ESQDRQHLH	878
Qy	951	PWARTPRLSVMEEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTKVYKDDKKLALNK	1010
Db	879	PDQKGP-----SEVC-----KEEPEMHLSSLGVSIQNLVKVYRDGKKVAVDG	919
Qy	1011	LSLNLYENQVVSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGM	1070
Db	920	LTLNFYEGQITSFLGHNGAGKTTTMSILTGLFPPTSGTAFILGKDIRSELSTIRQNLGVC	979
Qy	1071	PQHNVLFDRLTVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSVLVQTLSSGGMKR	1130
Db	980	PQHNVLFDLLTVEEHIWFYARLKGLPEKKVKEEMEQMAMDVGLPHKLGKARTSKLSGGMQR	1039
Qy	1131	KLSVAIAFVGGSRRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGD	1190
Db	1040	KLSVALAFVGGSKVILDEPTAGVDPYSRRGIWELLKLYRQGRTIILSTHHMDEADILGD	1099
Qy	1191	RIAIISHGKLKCCGSPLFLKGTGYDGYRLTLVKRPAEPG-----	1229
Db	1100	RIAIISHGKLCCVGSFLKNQLGTGYLTLVKKDVDSLSLSSCRNSSSTVSYLKKDDSVS	1159
Qy	1230	-GPQEPGLASSPPGRAPLSSCSELQVSQFIRKHAVASCLLVSDTSTELSYILPSEAAKKGA	1288
Db	1160	QSSSDAGLGSDDHESDTLTIDVS--AISNLITKHVPEARLVEDIGHELTYVLPYKAAKEGA	1217
Qy	1289	FERLFQHLERSLDALHLSSFGLMDTTLEEVFLKVSEEDQSLENSEADVKEERKDVLPGAE	1348
Db	1218	FVELFHEIDDRSLDLGISSYGISETTLEEIFLKVADD-----SGVDA-ETSDGTLP---	1267
Qy	1349	GPASGEGHAGNLARCSELTQSQASLQSASSVGSARGDEGAGYTDVYGD----YRPLED--	1402
Db	1268	-----ARRNRRA-----FGDRQSCSLRPFTEDD	1289
Qy	1403	--NPQDPDNVSLQEVEAEALSRV-GQGSRKLDGGWLKVRQFHGLLVKRFHCARRNSKALF	1459
Db	1290	AFDPNDSD-IDPESRETDLISGMDGKGSYQMGWKLSSQQQFMALLWKRLLIAKRSRKGFF	1348
Qy	1460	SQILLPAFFVCVAMTVALSVPEIGDLPLPLVLSPSQY-HNYTQPRGNFIPYANEERREYRL	1518
Db	1349	AQIVLPAVFVCIALMFSLIVPPFGKYPSLELQPMWYDEQYT-----FISNDAPE---	1397
Qy	1519	RLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFDSMCLE	1578
Db	1398	-----DAGTQKLLDALLNKPGFGTRCM-----	1419

Qy	1579	SFTQGLPLSNFVPPPPSPAPSDSPASPEDLQAWNVSLLPPTAGPEMWTSAFSLPRLVREP	1638
Db	1420	---QGHSI-----PD-----TPCTVQGKEWTTA-SVPDSVLEI	1448
Qy	1639	VR-----CTCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILTITGHNVSEYL	1683
Db	1449	LRGNWSMENPSPSPCECSNEKIKKMLPVCPPGAGGLPPPQREQDTADILQNLTGRIISDYI	1508
Qy	1684	LFTSDRF-----RLHRYGAITFG---NVLKSIPASFGTRAPPMVRKIA-----	1723
Db	1509	VKTYAQIIGKSLKNKIWVNEFRYGGFSLGARSSHVLP--PSNEVTDATKQVKKILELAQG	1566
Qy	1724	-----VRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAA	1765
Db	1567	SSGDRFLNNLASFMKGLDTKNNVKVWFNNKGWHAIASFLNVINNAILRANLQQGK-NPSA	1625
Qy	1766	YGITVTNHMPNKTSAFSL-LDYLLQGTDVVIAIFIIVAMSFVPASFVFLVAEKSTKAKH	1824
Db	1626	YGITAFNHPLNLTQQLSEVALMTTSVDVLVVICVIFAMSFVPASFVFLIQERVSKAKH	1685
Qy	1825	LQFVSGCNPIIYWLANYVWDMNLNVLVPATCCVILFVFDLPAYTSPTNFPVAVLSLFLLYG	1884
Db	1686	LQFISGVKPVIIYWLANFVWDMCNIIYVPATLVIIIFICFQQKSYVSSSNLPVLALLLLLYG	1745
Qy	1885	WSITPIMYPASFVFEVPSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVVNSYLKS	1944
Db	1746	WSITPLMYPASFVFKIPSTAYVVLTSVNLFIGINGSVATFVLELFTNNK-LNNINDILKS	1804
Qy	1945	CFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGF	2004
Db	1805	VFLIFPHFCLGRGLIDMVKNQAMADALERFGE-NRFVSPLSWDLVGRNLFAMAVEGVVFF	1863
Qy	2005	LLTIMCQYNFLRRPQRMVPSTKPVED-DVDVASERQVRVLRGDADNDMVKIENTLTKVYKSR	2063
Db	1864	LITVLIQYRFFIKPRPVYAKLPVNDDEDVNRERQRIISGGGQSDILEIRELTKIYRMK	1923
Qy	2064	KIGRILAVDRLCLGVRPGECFGLLGVNGAGKTSTFKMLTGDESTTGGEAFVNGHSLVKEL	2123
Db	1924	---RKPAVDRICVGIPPGECFGLLGVNGAGKSSTFKMLTGDTDVTGGDAFLKGNLSILNI	1980
Qy	2124	LQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKP	2183
Db	1981	QEVHQNMGYCPQFDVAVNELLTGREHLEFFALLRGVPEKEVCKVGEWAIRKGLVKYGEKY	2040
Qy	2184	AGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIKTRGSVVLTS	2243
Db	2041	AGNYSGGNRRKLSTAIALIGGPPVVFLEPTTGMDPKARRFLWNCALSVIKEGRSVVLTS	2100
Qy	2244	HSMECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVR-TKSSQSVKDVVRFFNR	2302
Db	2101	HSMECEALCTRMAMVNGRFRCLGSGVQHLKNRFGDGYTIVVRIAGGNPDLKPVEEFFGH	2160
Qy	2303	NFPEAMLERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFVNFA	2362
Db	2161	AFPGSVLKEKHRNMLQYQLPSSQSSLARIFSVLSQNKRLHIEDYSVSQTTLDQVFNFA	2220
Qy	2363	KKQSDN	2368



Db 2221 KDQSDD 2226

RESULT 5

Q80ZB2

ID Q80ZB2 PRELIMINARY; PRT; 2201 AA.  
AC Q80ZB2;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE ATP-binding cassette 1.  
GN ABCA1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley;  
RA Ananthanarayanan M., Mirza M.F.;  
RT "Cloning and Characterization of Rat Liver Abcal.";  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY208182; AAO53557.1; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
DR GO; GO:0000166; F:nucleotide binding; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR Pfam; PF00005; ABC\_tran; 2.  
DR ProDom; PD000006; ABC\_transporter; 2.  
DR SMART; SM00382; AAA; 2.  
DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
KW ATP-binding.  
SQ SEQUENCE 2201 AA; 246553 MW; B1472978BFC3E6B8 CRC64;

Query Match 32.2%; Score 4083; DB 11; Length 2201;  
Best Local Similarity 38.9%; Pred. No. 1.5e-261;  
Matches 969; Conservative 332; Mismatches 684; Indels 504; Gaps 63;

QY 62 LTSAGILPVMQSLCPDQORDEFGFL-----QYANSTVTQLLERLDRVVEEGLNFD 111  
: ||| || : | : : | : : | | : | : | :  
Db 1 MPSAGTLPWVQGIICNANNPCFRYPTPGEAPGVVGNFNKSIVSRLESDAQRIL----LYS 56  
  
QY 112 PARPSLGSELEALR--QHLEALSAGPGTSGSHLDRSTVSSFSLDVARNPQELWRFLTQN 169  
| : : || | : : : : | | | | :  
Db 57 QKDTSIKDMHKVLRITLQKIKHPNSNLKLQDFLVDNETFSG-----FLQHS 101  
  
QY 170 LSLPNSTAQALLAARVDPPEVY-----HLLFGPSSALDSQSGLHKGQEPWSRLGGNPLF 223  
||| | || | : : || : | |  
Db 102 LSLPRSAVDNLLQADVSLQKVFLQGYQLHL-----ASLCNGS----- 138  
  
QY 224 RMEELLAPALLEQLTCTPGSGELGRILTVPESQKGAQGYRDAVCSGQAAARARRFSG 283  
: || : : || | : | : | : :

Db	139	KLEELIR-----PEDLKVS-----ALCS----LPREKLDAP	165
Qy	284	SAELRNQLDVAK-VSQQGLGLDAPNGSDSSPQAPPPRRQLALLGLDLDA-----	330
Db	166	ERELRSNMIDILKPVMTKL-----NSTSLLPTQHLAEATTLLDSLGLLAQELFSTK	216
Qy	331	-----QKVL-----QDQDVLSALALLLPQGACTGRTPGPPASGAGGAA-----	368
Db	217	SWSDMRQEVMTLNTVNNSGSSSTQIYQAVSRIVCGHPEG-----GGLKIKSL	262
Qy	369	---NGTGAGAVMGPNTAEAGAPSAAALATP---DTLQGQCSAFVQ---LWAGLQPILCGN	420
Db	263	NWYEDNNYKALFGGNGTEEDTDTFYDNSTTPYCNDLMKNLESSPLSRIIWKALKPLLIG	321
Qy	421	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF	480
Db	322	-----KILYTPDTPATRQVMAEVNKT	343
Qy	481	AFVGNVTHYAQVWLNISAEIRSFLEQG-----RLQQHRLW-LQQY	519
Db	344	QELALFPDLEGMWEELSPOIWTFMESSQEMDLVRPMLDLRGNDQFWERKLDGLYWTAQDI	403
Qy	520	VAELRLHPEAL---NLSLDELPPALRQDNFSLPSGMALLQQLDITDNAACGWIQFMSKVS	576
Db	404	MAFLAKNPEDVQSPNGSVYTWREAFNETN-----QAIQTIS-----RFMECVN	446
Qy	577	VDIFKGFDPDEESIVNYTLNQAYQDNVTVFASVIFQ---TRKDGSLLPHVHYKIRQNSSFTE	634
Db	447	LNKLEPIPTTEVTLINKSMD---LLDARKFSAGIDFTGITPDSVELPHHVKDKIRMDIDNVE	504
Qy	635	KTNEIRRAYWRPGPNTG---GRFYFLYGFVWIQDMMERAIIDTFVGHDVVEPGSYVQMFP	691
Db	505	RTNKIKDGYWDPGPRADPFEDMRYVWGGFAYLQDVVEQAIIRVLTGTE-KKTGVYVQQMP	563
Qy	692	YPCYTRDDFLFVIEHMMPLCMVISWVYSVAMTIQHIVA EKEHRLKEVMKTMGLNNAVHWV	751
Db	564	YPCYVDDIFLRVMSRSMPLFMTLAWIYSVAVIKSI VYEKEARLKETMRIMGLDNGILWF	623
Qy	752	AWFITGFVQLSISVTALTAILKYQVLMHSHVVIWLFLAVYAVATIMFCFLVSVLYSKA	811
Db	624	SWFISSLIPLLVSAGLLVIIKLGLDLLPYSDPSVVFVFLSVFAVVTILQCFLISTLFSRX	683
Qy	812	KLASACGGIIFYLSYVPYMYVAIREEVAHDKITAFE-KCIASLMSTTAFGLGSKYFALYE	870
Db	684	NLAAACGGIIFYTLPLPYVLC-----VAWQDYVGFSIKIFASLLSPVAFGFGCEYFALFE	738
Qy	871	VAGVGIQWHTFSQSPVEGDDFNLLAVTMLMVDVAVVYGILTWYIEAVHPGMYGLPRPWYF	930
Db	739	EQGIGVQWDLNLFKSPVEEDGFNLTTSVSMMLFDTFIYGVMTWYIEAVFPGQYGIPRPWYF	798
Qy	931	PLQKSYWLGSGRTEAWESWPWARTPRLSVMEEDQACAMESRRFEETRGMEEEPTHLPLV	990
Db	799	PCTKSYWFGE---EIDEKSHPGS-----SQKGASEIC-----MEEEPHTHLKLG	838
Qy	991	VCVDKLTKVYKDDKKLALNKLNLNYENQVVSFLGHNGAGKTTTMSILTGLFPPTSGSAT	1050
Db	839	VSIQNLVKVYRDGMKVAVDGLALNFYEGQITSFLGHNGAGKTTTMSILTGLFPPTSGTAY	898

Qy	1051	IYGHDIRTEMDEIRKNLGMCPOQHNVLFDRLTVEEHLWFYSRLKSMAQEEIRREMDKMIED	1110
Db	899	ILGKDIRSEMNSIRQNLGVCPQHNVLFDMLTVEEHIWFYARLKGLSEKHVKAEMEQMALD	958
Qy	1111	LEL-SNKRHSLVQTLSGGMKRKLSVAIAFVGGSRAIILDEPTAGVDPYARRAIWDLILKY	1169
Db	959	VGLPPSKLKSQTSQLSGGMQRKLSVALAFVGGSKVVLIDEPYARRGIWELLKY	1018
Qy	1170	KPGRTIILLSTHHMDEADLLGDRIAIISHGKLKCCGSPLFLKGTGYRDLTVKRPAEPG	1229
Db	1019	RQGRTIILLSTHHMDEADILGDRIAIISHGKLCCVSSSFLKQLGTGYLTLVKKDVESS	1078
Qy	1230	-----GPQEPGLASSPPGRAPLSSCSELQVSQFIRKHAVASCLL	1267
Db	1079	LSSCRNSSSTVSLCKKEDSVSQSSSDAGLGSDESHTLTIDVS--AISNLIRKHVSEARL	1136
Qy	1268	VSDTSTELSYILPSEAAKGAERLRFQHLERSLDALHLSSFGLMDTTLEEVFLKVSEEDQ	1327
Db	1137	VEDIGHELTYVLPYEAKEGAFVELFHEIDRLSDLGISYGISETTLEEIFLKVAEE--	1194
Qy	1328	SLENSEADVKESSKDVLPGAEGPASGEGHAGNLARCELTQSQASLQSASSVGSARGDEG	1387
Db	1195	----SGVDA-ETSDGTLP-----ARRNRR	1213
Qy	1388	AGYTDVYGDYRPLF-----DNPQDPDNVSL--QEVEAEALSRLV--GQSRKLDGGWLKVRQ	1439
Db	1214	A-----FGDKQSCSLHPFTEDDAVDPNDSLDLPESRETDLGSGMDGKGSYQLKGWKLTTQQQ	1268
Qy	1440	FHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPPLVLSPSQYH-NY	1498
Db	1269	FVALLWKRLLIARRSRKGFFAQIVLPAVFVCIALVFSLVPPFGKYPNLELQPWMYNEQY	1328
Qy	1499	TQPRGNFIPYANEERREYRLRLSPDASPPQLVSTFRLPSGVGATCVLKSPANGSLGPTLN	1558
Db	1329	T-----FVSNDAP-----DMGTQELLNALTCKDPGFGRM-----	1359
Qy	1559	LSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDLDLQAWNVSLLP	1618
Db	1360	-----EGNPIPN-----TPC	1369
Qy	1619	TAGPEMWTSAAPSLPRLVREPVR-----CTCSAQGTGFS---CPSSVGG-HPPQ	1662
Db	1370	LVGEEDWTTGP-VPQTLMDLFQNGNWTMKNPSPSCQCSSDKIKMLPVCPPGAGGLPPPQ	1428
Qy	1663	MRVVTGDILTITGHNVSEYLLFTSDRF-----RLHRYGAITFG-----	1701
Db	1429	RKQKTADILQNLTKGRNNSDYLVKTYVQIIAKSLKNKVWVNEFRYGGFSLGVSDSQALPPS	1488
Qy	1702	---NVLKSIP-----ASFGRAPPMVRKIAVRRAAQVFYNNKGYHSMPT	1742
Db	1489	QEVNNAIKQMKLLKLTCDSSADRFLSSLGR---FMTGLDTKNNVKVWFNNKGWHAISS	1544
Qy	1743	YLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLS-LDYLLQGTDVVIAIFIIV	1801
Db	1545	FLNVINNAILRANLQGE-NPSOYGITAFNHPLNLTKOOLSEVALMTTSVDVLVVICVIF	1603



RX MEDLINE=97345663; PubMed=9202155;  
 RA Azarian S.M., Travis G.H.;  
 RT "The photoreceptor rim protein is an ABC transporter encoded by the  
 RT gene for recessive Stargardt's disease (ABCR).";  
 RL FEBS Lett. 409:247-252(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6;  
 RA Azarian S.M., Travis G.H.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF000149; AAC23916.1; -.  
 DR MGD; MGI:109424; Abca4.  
 DR GO; GO:0005548; F:phospholipid transporter activity; IMP.  
 DR GO; GO:0006649; P:phospholipid transfer to membrane; IMP.  
 DR GO; GO:0007601; P:vision; IMP.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR005951; Rim\_ABC\_transpt.  
 DR Pfam; PF00005; ABC\_tran; 2.  
 DR ProDom; PD000006; ABC\_transporter; 2.  
 DR SMART; SM00382; AAA; 2.  
 DR TIGRFAMs; TIGR01257; rim\_protein; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
 KW ATP-binding.  
 SQ SEQUENCE 2310 AA; 260207 MW; 8370C6C8A62EF294 CRC64;

Query Match 31.2%; Score 3951; DB 11; Length 2310;  
 Best Local Similarity 36.4%; Pred. No. 9.7e-253;  
 Matches 942; Conservative 396; Mismatches 771; Indels 480; Gaps 65;

Qy	1	MGFLHQLQLLLWKNVTLKRRSPWVLA	FEIFIPLVLFFILLGLRQKKPTISVKEVPFYTAA	60
		:          ::	:        :	
Db	1	MGFLRQIQLLLWKNWTLRKRQKIRFVVELVWPLSLFLVLIWLRNANPLYSQHECHFPNKA		60
Qy	61	PLTSAGILPVMQSL-----C---PDGQRDEF	GFLQYANSTVTQLLERLDRVVEEGNLF	110
		:    :    :   :	:    :   :	
Db	61	-MPSAGLLPWLQGIFCNMNPFCFQNPTPGESPGTVSNYNNS----	ILARVYRDFQELFMD	115
Qy	111	DPARPSLG---SELEALRQHLEALSAGPGTSGSHLDRSTVSSFS	SLDSVARNPQELWRFLT	167
		:      ::	:    : : :	
Db	116	TPEVQHGLGQVWAE LR T L S Q F M D T L R -----	THPERFAGRGLQIRDILKDEEALTFLM	168
Qy	168	QNLSPNSTAQALLAARVDPPEVYHLLFGPSSALDSQSG	LHKQEPWSRLGGNPLFRMEE	227
		: :   :      : ::  :	: :	
Db	169	RNIGLSDSVAHLLVNSQVRVEQFAY-----	GVPDELETD	202
Qy	228	LLLAPALLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSG	QAAARARRFSGLSAEL	287
		: :    ::	:   :      :  :	
Db	203	IACSEALLQRF-----	IIFSQRGAQTVRDALCPLSQVT----LQWIEDTL	244
Qy	288	RNQLDVAKVSQQGLDAPNGSDSSPQAPPPRRQLALLGDLLD	-AQKVLQDQVDVLSALALL	346
		:    :	:        :     :	
Db	245	YADVDFFKLFHVL----	PTLLDSSSQGINLRFWGGILSDLSPRMQKFIHRPSVQDLLWVS	300
Qy	347	LPQGACTGRTPGPPASGAGGAANGTGAGAVMGP	NATAEEGAPSAAALATPDTLQGQCSAF	406
			:	

Db	301	RP-----LLQNGGP-----ETF	312
Qy	407	VQLWAGLQPILCGNNRTIEPEALRRGNMSSLGFTSKEQRN----LGL-----	449
Db	313	TQLMSILSDLLCG-----YPEG---GGSRVFSFNWYEDNNYKAFLGIDSTRKDPAYSYDK	364
Qy	450	-----LVHLMTSNP-----KILYAPAGSEVDRVILKANETFAFVGNV	486
Db	365	RTTSFCNSLIQSLESNPLTKIAWRAAKPLLMGKILFTPDSPAARRIMKNANSTFEELDRV	424
Qy	487	THYAQVWLNISAEIRSFLEQGRLOQHLR-----WLQQYVAELRLHPEA-LNL---	532
Db	425	RKLVKAWEEVGPQIWIYFFFEKSTQMTVI RD TLQHPTVKDFINRQLGEEGITTEAVLNFFSN	484
Qy	533	-----SLDELPPALRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEE	587
Db	485	GPQEQKQADDMTSFDWRDIFNITDRFLRLAN-----QYLECLVLDDKFESYDDEV	532
Qy	588	SIVNYTLNQAYQDNVTVFASVIFQTRKD--GSLPPHVHYKIRQNSSFTEKTNEIRRAYWR	645
Db	533	QLTQRALSLLEENR--FWAGVVFPGMPWASSLPPHVKYKIRMDIDVVEKTNKIKDRYWD	590
Qy	646	PGPNTGGREYFLY---GFVWIQDMMERAIIDTFVGHDVVEP--GSYVQMFPYPCYTRDDF	700
Db	591	SGPRADPVEDFRYIWWGGFAYLQDMVEQGIVKSQM--QAEPPIGVYLQOMPYPFCVDDSF	647
Qy	701	LFVIEHMMPLCMVISWVYSVAMTIQIHIVAEKEHRLKEVMKTMGLNNAVHWVAVFITGFVQ	760
Db	648	MIILNRCFPIFMVLAWIYSVSMTVKGIVLEKELRLKETLKNQGVSSNAVIWCTWFLDSESI	707
Qy	761	LSISVTALTALIKYGQVLMHSHVVIWFLAVYAVATIMFCFLVSVLYSKAKLASACGGI	820
Db	708	MALSIFLLTLFIMHGRILHYSDFILFLFLAFATATIMQSFLSTLFSKASLAAACSGV	767
Qy	821	IYFLSYVPYMYVAIREEVAHDKITAFEKCIASLMSTTAFGLGSKYFALYEAVGVGIQWHT	880
Db	768	IYFTLYLPHVLCFAWQ---DRMTADLKTTVSLLSSVAFGFGTEYLVRFEQGLGLQWSN	823
Qy	881	FSQSPVEGDDFNLLLAVTMLMVDVAVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGS	940
Db	824	IGKSPLEGDEFSSFLLSMKMMLLDAALYGLLAWYLDQVFPDYGTPLPWYFLLQESYWLGG	883
Qy	941	GRTEAWESWPWARTPRLSVMEEDQACA-MESRRFEETRGMEEE---PTH-----	986
Db	884	-----EGCSTREERALEKTEPLTEEMEDPEHPEGMNSFFFE	919
Qy	987	--LP-LV--VCVDKLTkVYKDDKKLALNKLNLNLYENQVVSFLGHNGAGKTTMSILTGL	1041
Db	920	RELPLVPGVCVKNLVKVFEPSGRPVDRNLNITFYENQITAFGLHNGAGKTTLSILTGL	979
Qy	1042	FPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSRLKSMAQEEIR	1101
Db	980	LPPTSGTVLIGGKDIETNLDVVRQSLGMCPQHNLFHHLTVAEHILFYAQLKGRSWEEAQ	1039
Qy	1102	REMDKMIEDLELSNKRHSLVQTLSGMKRKLSSVAIAFVGGSRRAILDEPTAGVDPYARRA	1161
Db	1040	LEMEAMLEDTLGHHKRNEEAQDLSGMORKLSVAIAFVGDSKVVVLDEPTSGVDPYSRRS	1099

Qy	1162	IWDLILKYPGRTILLSTTHMDEADLLGDRIAIISHGLKCCGSPLFLKGTYGDBGYRLTL	1221
Db	1100	:    :     :              :        :	
Qy	1222	VKR----PAEPGGPQ-----EPLGLASSPPGRAPLSSCSELQV-----SQFIRKHVA	1263
Db	1160	::        ::    :              ::     :        :                     :	
Qy	1264	SCLLVSDTSTELSYILPSEAAKKGAFERLFQHLEERSLDALHLSSFGMLMDTTLEEVLKVS	1323
Db	1218	::  ::   :   :   :    :         :       :    :	
Qy	1324	EEDQSLENSEADVKE SRKDVLPGAEGPASGEHAGNLARCELTQS QASLQSASSVGSAR	1383
Db	1278	: :                ::   :                                   :   :	
Qy	1384	GDEGAGYT DVYG DY RPLFDNPQDPDNVSLQEVEAEALS RVGQSRKLDGGWLKVRQFHGL	1443
Db	1327	:                  :                         :                                             : :	
Qy	1444	LVKRFHCARRNSKALFSQILLPAFFVCVAMTVALS VPEIGDL PPLVLSPSQY-HNYT---	1499
Db	1358	:          ::       :  : ::             :	
Qy	1500	--QPRGNFIPIYANEERREYRLRLSPDASPQQIVSTFRLP SGVGATCVLKSPANGSLGPTL	1557
Db	1418	:                              :	
Qy	1558	NLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPEDLQAWNVS LP	1617
Db	1439	:   :        :                   :                             ::	
Qy	1618	PTAGPEMWTSAPSLR LRVREPVRCTCSAQGTGF---SCPSSVGGHP PQMRVV-TGDILT D,	1673
Db	1471	:   :                               :                                         : ::	
Qy	1674	ITGHNVSEYLLFT-----SDRFR LH--RYGAITFG NVLSIPAS-----	1710
Db	1524	:          : : :                                          :  ::              :            :	
Qy	1711	--FGTRAPPMVRK-----IAVRRAAQVFYNNGYHSMP TYLN SLNNAILRANLP	1757
Db	1584	:   :                :                : :      :  : :           :        :	
Qy	1758	KSKGNPAAYGITVTNHPMNKTSASLS-LDYLLQGT D VVIAIFII VAMS FVPASFVV FLVA	1816
Db	1644	: : :                :   :             :            :   :                  :  :	
Qy	1817	EKSTKAKHLQFVSGCNPIIYWLANYVWDM L NYLVPATCCVII LFVFDLPAYTSPTNFPAV	1876
Db	1703	:                :              : :  :                                           :	
Qy	1877	LSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLK	1936
Db	1763	:   :    : :          :             :                    : :          :   :    : :	
Qy	1936	VSLLMLYGWAVIPMY PASFLFEVPSTAYVALSCANIFIGINSSAITFVLELFENNRTL	1822
Db	1763	:    : :          :             :                    : :          :   :    : :	





RL J. Biol. Chem. 272:10303-10310(1997).  
 DR EMBL; U90126; AAC48716.1; -.  
 DR GO; GO:0005887; C:integral to plasma membrane; IEA.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR005951; Rim\_ABC\_transpt.  
 DR Pfam; PF00005; ABC\_tran; 2.  
 DR ProDom; PD000006; ABC\_transporter; 2.  
 DR SMART; SM00382; AAA; 2.  
 DR TIGRFAMs; TIGR01257; rim\_protein; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
 KW ATP-binding.  
 SQ SEQUENCE 2281 AA; 257228 MW; 71CD404C98F7A079 CRC64;

Query Match 30.7%; Score 3893.5; DB 6; Length 2281;  
 Best Local Similarity 36.3%; Pred. No. 6.3e-249;  
 Matches 937; Conservative 390; Mismatches 772; Indels 483; Gaps 70;

Qy	1	MGFLHQLQLLLWKNVTLKRRSPWVLA FEIFIPLVLF FILLGLRQK KPTISVKEVPFYTAA	60
		::          ::    :       :  :	
Db	1	MGFARQIKLLWKNWTLRKRQKIRFVVELVWPLSLFLVLIWLRNVNPLYSKHECHFPNKA	60
Qy	61	PLTSAGILPVMQSL-C----PDGQRDEFG-----FLQYANSTVTQLLERLDRVVEEGNLF	110
		:    :    :  :                     :  :  :  :	
Db	61	-MPSAGMLPWLQGIFCNVNNPCFQSP TAGESPGIVSNYNNS----ILARVYRDFQELMD	115
Qy	111	DPARPSLGSELEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNL	170
		:  :   : :      :  :	
Db	116	APESQHLGQVWREL R----TLSQLMNTLRMHPERIAGRGIRIREVLKDDMLTLFLVKNI	171
Qy	171	SLPNSTAQALLAARVDPPEVYHLLFGPSSALDSQSGLHKQEPWSRLGGNPLFRMEELL	230
		:    : :      :        :  :  :	
Db	172	GLSDSVVYLLVNSQVRP-----EQFAR--GVPDLMLKDIAC	205
Qy	231	APALLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAELRNQ	290
		:              :      :  :  :  :	
Db	206	SEALLE-----RFLIFP--QRRAAQTVRGSLSLSQGT----LQWMEDTLYAN	247
Qy	291	LDVAKVSQQGLGDAPNGSDSSPQAPPPRRRLQALLGDLLD-AQKVLQDQVDVLSALALLPQ	349
		:    :            :    :  :      :  :	
Db	248	VDFFKLFHVF----PRLLDSRSQGMNLRSWGRI LSDMSPRIQEFIH RPSVQDLLWVTRP-	302
Qy	350	GACTGRTPGPPASGAGGAANGTGAGAVMGP NATAE EGAPSAAALATPDTLQGQCSAFVQL	409
		:	
Db	303	-----LVQTGGP-----ETFTQL	315
Qy	410	WAGLQPILCGNNRTIEPEALRRGNMSSLGFTSKEQRN----LGL-----	449
		:                       :	
Db	316	MGILSDLLCG----YPEG---GGSRVFSFNWYEDNNYKAFLGIDSTRKDPIYSYDERTT	367
Qy	450	-----LVHLM TSNP-----KILYAPAGSEVDRVILKANETFAFVGNVTHY	489
		: :         :     :  :       :	

Db 368 TFCNALIQSLESNPLTKIAWRAAKPLLMGKILFTPDSPATRRILKNANSTFEELERVRL 427

Qy 490 AQVWLNISAEIRSFLEQGRLQOHLR-----WlQQYVAELRLHPEALNLSLDELP 538  
:|| : :| | : : :| :| : | : ||: | : |

Db 428 VKVWEEVGPQIWYFFDKSTQMSMIRDITLENPTVKAFWNRQ-LGEEGITAEAV---LNFLY 483

Qy 539 PALRQDNFSLPSGMALLQQLDTIDNAACGW-----IQFMSKVSVDIFKGFP 584  
| : | | :|| | :| : :| | : :

Db 484 NGPREG-----QADDVDN--FNWRDIFNITDRALRLANQYLECLILDKFESYD 529

Qy 585 DEESIVNYTLNQAYQDNVTVFASVIFQTRK--DGS�PHVHYKIRQNSSFTEKTNEIRRA 642  
|| : | : : :| | :| | | | | | : | | | :| :

Db 530 DEFQLTQRALSLLLEENR--FWAGVVPDMPHTSSLPHPVKYKIRMDIDVVEKTNKIKDR 587

Qy 643 YWRPGPNTGGRFYFLY---GFVWIQDMMERAIIDTFVGHDDVEPGSYVQMFYPYCYTRDD 699  
|| || | | || :||| : | : : | | | : | | | : |

Db 588 YWDSGPRADPVEDFRYIWGGFAYLQDMVEHGITRS-QAQEEVPVGIYLOQMPYPCFVDDS 646

Qy 700 FLFVIEHMMPLCMVISWVYSVAMTIQHIVAKEHRLKEVMKTMGLNNAVHVVAVFITGFV 759  
| : : : | : || : || : || : || : || | | | | | : | | : | | | | | :

Db 647 FMIILNRCFPIFMVLAWIYSVSMTVKISIVLEKELRLKETLKNQGVSNRVIWCTWFLDSEFS 706

Qy 760 QLSISVTALTALIKYGQVLMHSHVVIWFLAVYAVATIMFCFLVSVLYSKAKLASACGG 819  
:| :| : || : :| :| : | :|| : :||| || :| :| :| | | | |

Db 707 IMSMSICLLTIFIMHGRILHYSNPFILFLLAFSIATIMQCFLSTFFSRASLAAACSG 766

Qy 820 IIFYLSYVPYMYVAIREEVAHDKITAFEKCIASLMSTTAFGLGSKYFALYEVAGVGIQWH 879  
:||| | :| : : : | :|| | | | :| | | | :| :| | | :| |

Db 767 VIYFTLYLPHILCFWQ----DRITADMKMAVSLSPVAFGFGTEYLAXFEEQGVGLQWS 822

Qy 880 TFSQSPVEGDDFNLLLAVTMLMVDVAVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLG 939  
|| :||| :| : :| : :| :| :| | | | | | | | | | | | | | | |

Db 823 NIGNSPMEGDEFSLMSMKMMLLDAALYGLLAWYLDQVFPDYGTPLPWYFLLQESYWLG 882

Qy 940 SGRTEAWESWPPWARTPRLSVMEEDQACA-MESRRFEETRGMEEE----- 983  
: | : | | | :| : | |

Db 883 G-----EGCSTREERALEKTEPITEEMEDPEYPEGINDCFF 918

Qy 984 PTHLP-LV--VCVDKLTKVYKDDKKLALNKLNLNLYENQVVSFLGHNGAGKTTTMSILTG 1040  
|| || || | | : : : | : :| : : || : : | | | | | | | | | | | |

Db 919 ERELPGLVPGVCVKNLVKIFEPYGRPAVDRLNITFYESQITAFGLGHNGAGKTTTLSIMTG 978

Qy 1041 LFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSRLKSMAQEEI 1100  
| | | | : : | | | : | | :| | | | | | | | | | | | | | | | | | |

Db 979 LLPPTSGTVLVGGKDIETNLDAIRQSLGMCPQHNLFHHLTVAEHILFYAQLKGRSWDEA 1038

Qy 1101 RREMDKMIEDLELSNKRHSLVQTLSGGMKRLSVAIAFVGGSRAILDEPTAGVDPYARR 1160  
: || :| :| | :| : : | | :| | | | | | | | | | : : | | | | | | | |

Db 1039 QLEMEAMLEDTLGHHKRNEEARDLSGGVQRKLSVAIAFVGDAKVVVLDEPTSGVDPYSRR 1098

Qy 1161 AIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPLFLKGTYGDGYRLT 1220  
:| | :| :| : | | :| | | | | | | | | | | | | | | | | | | | | | |

Db 1099 SIWDLLLKYRSGRTIIMSTHHMDEADILGDRIAIISQGRLYCSGTPLFLKNCFGTGFYLT 1158

Qy 1221 LVKRP---AEPGGPQEPGLASSPPG---RAPLSSCSEL-----QVSQFIRKHV 1262  
|| :| | :| : : | | | :| : | : : : : |

Db 1159 LVRRMKTIQSQGRGREATCSCASKGFSVRCP--ACAEAITPEQVLDGDVNELTDMVHHHV 1216

Qy	1263	ASCLLVSDTSTELSYILPSEAAKKGAFLRFQHLESLDALHLSSFGMLMDTTLEEVLKV	1322
Db	1217	PEAKLVEICIGQELIFLLPNKNFKQRAYASLFRELEETLADLGLSSFGISDTPLEEIFLKV	1276
Qy	1323	SEEDQSLENSEADVKEARKDVLPGAEGPASGEGHAGNLARCELTQSQASLQSASSVGSA	1382
Db	1277	TEDLDSGHLFAGGTQQKRENI--NLRHPCSG-----PSEKAGQTPQGSSSH	1320
Qy	1383	RGDEGAGYTDVYGDIRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQFHG	1442
Db	1321	PGEPA-----HPEGQPPP-----EREGHSRLNSGAR-----LIVQHVQA	1355
Qy	1443	LLVKRHFHCARRNSKALFSQILLPAFFVCVAMTVALSVP EIGDLPLVLSPSQY-HNYTQP	1501
Db	1356	LLVKRFQHTIRSHKDFLAQIVLPATFVFALMLSLIIPPFGEYPALT LHPWMYGQOYT--	1413
Qy	1502	RGNFIPIYANEERREYRLRLSPDASPQQVLVSTFRLP SGVGATCVLKSPANGSLGPTLNLS	1561
Db	1414	-----FFSMDQPDSEWLSAL-----ADVLVNKPG-----	1437
Qy	1562	GESRLLAAREFFDSMCL-ESFTQGLPLSNFVPPPPSPAPSDSPASPE---DIQAWNVS	1616
Db	1438	-----FGNRCLKEEWLPEFP CGN---SSPWKTPS---VSPDVTHLLQQQKW TADQ	1481
Qy	1617	PPTAGPEMWTSAPSLPRLVREPVRCTCSAQGTGF---SCPSSVG GHPQMRRVV-TGDILT	1672
Db	1482	P-----SPS-----CRCSTREKLTM LPECPEGAGGLPPPQRIQRSTEILQ	1521
Qy	1673	DITGHNVSEYLIFT-----SDRFR LH--RYGAITFG NVLKSIPAS-----	1710
Db	1522	DLTDRNVSDFLVKTYPALIRSSLKSKFWVNEQRYGGISVGG KLPAPPFTGEALVGFLSDL	1581
Qy	1711	---FGTRAPPMVRKIAVRRA-----QVFYNNKG YHSMPTYLNSLNNAILRANL	1756
Db	1582	GQLMNVS GGPMTREAAKEMPAFLKQLETEDNIK VWFNNKGWHALVSFLNVAHNAILRASL	1641
Qy	1757	PKSKGNPAAYGITVTNHMPMKTSASLS-LDYLLQGT DVVIAIFIIVAMSFVPASFVFLV	1815
Db	1642	HKDK-NPEEYGITVISQPLNLTKEQLSEITVL TTSDAVVAICVIFAMSFVPASFVLYLI	1700
Qy	1816	AEKSTKAKHLQFVSGCNPIIYWLAN YVWDMNLNYLVPATCCVII LFVFDLPAYTSPTNFPA	1875
Db	1701	QERVNKAKHLQFVSGVSPTTYWLTNFLWDIM NYTVS AALVVGIFIGFQKKAYTSSENLPA	1760
Qy	1876	VLSLFLLYGWSITPIMYPASF WFEPSSAYVFLIVINLFIGITAT VATFLLQLFEHDKDL	1935
Db	1761	LVALLMLYGWAVIPMMYPASF LFDIPSTAYVALSCANLFIGINSSAIT FVLELFENNRTL	1820
Qy	1936	KVVNSYLKSCFLIFPNYNLGHGLMEMAY NEYINEYYAKIQFDKM KSPFEWDIVTRGLVA	1995
Db	1821	LRINAMLRKLLIIIPPHFCLGR GLIDLALSQA VTDVYAQFGEAHS-SNPFQWDLIGNLAA	1879
Qy	1996	MAVEGVVGFLLTIMCQYNFL-----RRP QRM PVSTKPVEDDVDVASERQ RVLRGDADNDM	2050
Db	1880	MAVEGVVYFLTLTLLIOYOFFFSRW TTEPAKEPIT----DED DDVAERORIISGGNKTDI	1935

Qy 2051 VKIENLTKVYKSRKIGRILAVDRCLGVRPGECFGLLVNGAGKTSTFKMLTGDESTTGG 2110  
 ::: ||||| |||||:|||||:|||||:|||||: ||  
 Db 1936 LRLNELTKVYSGTSSP---AVDRLCVGVPRGECFGLLVNGAGKTTTFKMLTGDTAVTSG 1992

Qy 2111 EAFVNGHSLVKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWA 2170  
 :| | | |:| : | |:|||| | |: | | ||| | | |||: :: || |:  
 Db 1993 DATVAGKSILTNIISDVHQSMGYCPQFDAIDDLTGREHLYLYARLRGVPAEEIERVTNWS 2052

Qy 2171 LEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLIL 2230  
 :: | |: |||: |||||:||||| | : |||||:|||| | | |:  
 Db 2053 IQSLGLSLYADRLAGTYSGGNKRKLSTAIALIGCPPLVLLDEPTTGMDPQARRMLWNTIM 2112

Qy 2231 DLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVTRKSS 2290  
 :|: |:|||||:||||| | :|||:|||||:|||||:|::: |  
 Db 2113 GIIREERAVVLTSHSMEECEALCTRLAIMVKGAFCQLGTIQHLKSKFGDGYIVTMKIRSP 2172

Qy 2291 Q-----SVKDVVRFFNRNFPEAMLEKHHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIE 2345  
 : : | :|| ||| :: :|||: :|:| | |||:| : | ||  
 Db 2173 KDDLPLDLPVEQFFQGNFPGSVQRRHYNTLQFQVSSS---SLARIFRLLVSHKDSLLIE 2230

Qy 2346 DYSVSQTTLDNVFNFAKKQSDNLEQQETEPPSALQSPILGCLLSLLRPRSAPTELRALVA 2405  
 :|||:|||| | |||||:|:: : | ||:| :|  
 Db 2231 EYSVTQTTLQVFNFAKKQNETYDLP-----LHPRTAGASRQAKEV 2272

Qy 2406 DE 2407  
 |:  
 Db 2273 DK 2274

# RESULT 8

Q7TNJ2

ID Q7TNJ2 PRELIMINARY; PRT; 2170 AA.  
 AC Q7TNJ2;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE ATP-binding cassette transporter sub-family A member 7.  
 GN ABCA7.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wister; TISSUE=Platelet;  
 RA Sasaki M., Nada S., Yamaguchi A.;  
 RT "Cloning of rat ABCA7.";  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AB097814; BAC81426.1; -.  
 KW ATP-binding.  
 SQ SEQUENCE 2170 AA; 237718 MW; 003C8DF70B8744CE CRC64;

Query Match 29.1%; Score 3681; DB 11; Length 2170;  
 Best Local Similarity 35.8%; Pred. No. 7.8e-235;  
 Matches 911; Conservative 369; Mismatches 743; Indels 520; Gaps 71;

Qy 1 MGFLHQLQLLLWKNVTLKRRSPWVLAFEIFIPLVLFILLGLRQKKPTISVKEVPFYTAA 60

Db	1	MAFCTQLMLLLWKNYTYRRRQPIQLVVELLWPLFLFFILVAVRHSHPPLEHHECHF-PNK	59
Qy	61	PLTSAGILPVMQSLCPDGQRDEF-----GFL-QYANSTVTQLLERLDRVVEEGNLF	110
Db	60	PLPSAGTVPWLQGLVCNVNNSCFQHPTPGKEKPGVLSNFKDSLISRLLAHTVL-GGHST	118
Qy	111	DPARPSLGSELEALRQHLEALSAG--PGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQ	168
Db	119	QDMLAALGKLIPVLR----AVGSGAWPQESNQPAKQGSVT-----ELLEKILQ	162
Qy	169	NLSLPNSTAQALLAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEEL	228
Db	163	RASLETVLGQA-----QDSMRKFSDATRTVA-----QEL	191
Qy	229	LLAPALLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAELR	288
Db	192	LTLPSLV-----ELRALLRRPRGSAGSLELISEALCS-----	223
Qy	289	NQLDVAKVSQQGLDAPNGSDSSPQAPPFRRRLQALLGDLLDAQKVLQDQVDVLSALALLP	348
Db	224	-----	223
Qy	349	QGACTGRTPGPPASGAGGAANGTGAGAV---MGPNATAEEGAPSAAALATPD-TLQGQCS	404
Db	224	-----TKGPSSPG-GLSLNWYEANQINEFMGP-----ELAPT-----LPDSSLSPACS	265
Qy	405	AFV-----QLWAGLQPILCGNNRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLM	454
Db	266	EFVGALDDHPVSRLLRRLKPLILG-----	290
Qy	455	TSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVWLNISAEIRSFLEQ----GRLQ	510
Db	291	----KILFAPDTNFTRKLMQVNVQTFEELALLRDLHELWGVLPQIFNFMNDSTNVAMLQ	346
Qy	511	QHL-----RWLQQYVAELRLHPEALNLSLDELPPALRQDNF-SLPSGMALLQQLDTIDNA	564
Db	347	KLLDVEGTGW-QQOTPKGQKQLEAIR---DFLDPSRGRYNWQEAHADMGRLAEI-----	396
Qy	565	ACGWIQFMSKVSVDIFKGFPEESIVNYTLNQAYQDNVTVFASVIFQTRK---DGSLPP-	620
Db	397	-LG--QILECVSLDKLEAVPSEEALVSRALELLGERR--LWAGIVFLSPEHPLDSSEPPS	451
Qy	621	-----HVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLY---GFVWIQDMMERA	670
Db	452	PTTTGPGHLRVKIRMDIDDVTRTNKIRDKFWDPGPSADPLMDLRYVWGGFVYLQDLLEQA	511
Qy	671	IIDTFVGHDVVEPGSYVQMFYPYCYTRDDFLFVIEHMMPLCMVISWVYSVAMTIQHIVAE	730
Db	512	AVRVLSGRD-SRAGLYLQQMPHPCYVDDVFLRVLSRSLPLFLTALWIYSVALTVKAVVRE	570
Qy	731	KEHRLKEVMKTMGLNNAVHWVAFITGFVQLSISVTALTALIKYGQVLMHSHVVIWFL	790
Db	571	KETRLRETMRAMGLSRAVLWLGWFLSCLGPFVLSAALLVLVLKGNILPYSHPVVFLFL	630
Qy	791	AVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPY-MYVAIREEVAHDKITAFEKC	849

Db 631 AAFAVATVAQSFLLSAFFSRANLAAACGGLAYFALYLPYVLCVAWRERLPLGGLLA---- 686  
 Qy 850 IASLMSTTAFGLGSKYFALYEAVGVGIQWHTFSQSPVEGDDFNLLAVTMLMVDVAVYGI 909  
 ||:| ||| | : || | | || | | :| | :|||:|:  
 Db 687 -VSLSPVAFGFGCESLALLEEQGDGAQWHNLGTGPAE-DVFSLAQVSAFLLLDVAVIYGL 744  
 Qy 910 LTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSVMEEDQACAM 969  
 ||:||| || ||:| || || :||| | | | : ||: | |  
 Db 745 ALWYLEAVCPGQYGIPEPWNFPFRRSYWCGPG-----PPKSSVL----APAP 787  
 Qy 970 ESRRFEETRGMEEETHLPLVVCVDKLTKVYKDDKKLALNKLNLNLYENQVVSFLGHNGA 1029  
 : : :|| | | | : | | : : || | :| : || : :||| ||  
 Db 788 QDPKVL----VEEPPPGLVPGVSIRGLKKHFRGSPQPALRGLNLDYEGHITAFLGHNGA 843  
 Qy 1030 GKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWYF 1089  
 |||||:|||:|||||:||||:| |||:| | || :||:||||:|||| | |||||:|  
 Db 844 GKTTTLSILSGLFPSSSGSASILGHDVQTNMAAIRPHLGICPQYNVLFDMLTVEEHVWYF 903  
 Qy 1090 SRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVGGSRRAILDE 1149  
 ||| : : | | : :| | : | | : : |||||:|||||:||||| :|:|  
 Db 904 GRLLKGVSAAAIDSEQEHLIRDVGLIPKRDQTQRLHLSGGMQRKLSVAIAFVGGSRVIMDE 963  
 Qy 1150 PTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPLFL 1209  
 ||||| :|| ||:|:|:| : |||:|||||:|:|:|||||:|:| : | | |||||  
 Db 964 PTAGVDPASRRGIWELLLKYREGRTLILSTHHLDEAELLGDRVAMVASGSLCCCGSPLFL 1023  
 Qy 1210 KGTYGDGYRLTLVK-----RPAEPG-----GPQEPGLAS 1238  
 : | || |||| | : | |  
 Db 1024 RRHLGCGYYLTLVKSSQSLVTHDLKGDTEPRREKKSSEKGTADTVLTRDGPBRSSQVP 1083  
 Qy 1239 SPPGRAPLSSSELQVSQFIRKHVASCLLVSDTSTELSYILPSEAAKKGAFERLFQHLER 1298  
 :| | : : | : : :|| || : || || | | :| :| | :|  
 Db 1084 APDA-VPVTPSAAL-ILELVQRHVPGAQLVEELPHELVLALPYAGALDGSFATVFQELDQ 1141  
 Qy 1299 SLDALHLSSFGLMDTTLEEVFLKVSEEDQSLENSEADVKESTRKDVLPGAEGPASGEGHAG 1358  
 | : | | :| : || |||:|||| || : : | : : | | |  
 Db 1142 QLERLGLTGYGISDTNLEEIFLKVVEEAHA--HGEAGDPQQOHLTATPQP-----HTG 1194  
 Qy 1359 NLARCELTQSQASLQSASSVGSARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAE 1418  
 | | | : | : || | : || | : : :  
 Db 1195 PEASVLE-----NGELAKLVLDPAKGSAPTQAQVQ 1226  
 Qy 1419 ALSRVGQGSRKLDGGW-LKVRQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVAL 1477  
 || | :| || || |||: : |||:|:| || :| : |  
 Db 1227 -----GWT LTCQQLRALLHKRFLARRSRRGLFAQIVLPALFVGLALFFTL 1272  
 Qy 1478 SVPEIGDLPPLVLSPSQYHNYTQPRGNFIPYANEERREYRLRLSPDASPQQVLSTFRLPS 1537  
 || | ||| |||: | | : : :| : | : : :|  
 Db 1273 IVPPFGQYPPLQLSPAMY-----GPQVSFFSED-----APADPNRM----KLE 1312  
 Qy 1538 GVGATCVLKSPA---NGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPP 1594  
 : | : | : || | | || | | | : | | |  
 Db 1313 ALLGEAGLQDPSVQKGSRGSECTHS-----LACYFTVPEVPPDVASILASGNWTPDSP 1366  
 Qy 1595 SPAPSDSPASPDDELQAWNVSPLPTAGPEMWTSAPSLPRLVREPVRCTCSAQGTGF---S 1651  
 ||| | || |  
 Db 1367 SPA-----CQCSQPGARRLLPD 1383

Qy 1652 CPSSVGG-HPPQMRVVTGDILTDITGHNVSEYLLFTSDRFR LH-----RYGAIT 1699  
 ||: || ||| |:: ::|| |||::|: | ||| :  
 Db 1384 CPAGAGGPPPPQAMAGFGEVVQNLTGRNVSDFLVKTYPSLVRRGLKTKKWVDEVRYGGFS 1443

Qy 1700 FGNVLKSIPASFGTRAPPMVRKIA-----VRRAAQV 1730  
 | :|: :|| :| | : ::  
 Db 1444 LGGRDPDLPS-----GREVVRTVAEMRALLSPQPGNTLDRILNNLTQWALGLDARNSLKI 1498

Qy 1731 FYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLL-Q 1789  
 ::|||::|:| :| || :|| | | : || |||:| | || |:  
 Db 1499 WENNKGWHAMVAFVNRANGLLRAFLP-SGSVRHAHSITTLNHPLNLTKEQLSEATLIAS 1557

Qy 1790 GTDVVIAIFIIVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMLNYL 1849  
 ||:::| :: ||||| ||| : | : | :||| ||| :||| |::||| |||  
 Db 1558 SVDVLVSICVVFAMSFVPASFTLVLEERITRAKHLQLVSGLPQTLYWLGNFLWDMCNYL 1617

Qy 1850 VPATCCVIIILFV-FDLPAYTSPTNFPAVLSLFLLYGWSITPIMYPASFWFEPSSAYVFL 1908  
 | | | |:::|: | | :| | |::| | ||||| |||:|||||:| |||:||| |  
 Db 1618 V-AVCIVVLIFLAFQQKAYVAPENLPALLLLLLLYGWSITPLMYPASFFFSVPSTAYVVL 1676

Qy 1909 IVINLFIGITATVATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYIN 1968  
 ||||| |::|||:|:| |::|: | : || ||||:: || ||::| | : :  
 Db 1677 TCINLFIGINSSMATFVLELLS-DQNLQEVSRILKQVFLIFPHFCLGRGLIDMVRNQAMA 1735

Qy 1969 EYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPV 2028  
 : : ::| : :|| ||| : : |::| :| : |::: | : || | | |  
 Db 1736 DAFERLGD-KQFQSPLRWDIIGKNLLAMVAQGPIFLITILLQHRNRLLPQ--PKSRLPP 1792

Qy 2029 ---EDDVDVASERQVRVLRGDADNDMVKIENLTQVYKSRKIGRILAVDRLCLGVRPGECFG 2085  
 |:| || ||:|: | |:: : :|||: : | ||| |||: |||||  
 Db 1793 PLGEEDDVVRERERVTKGATQGDVLVLRDLTKVYRGQ---RSPAVDHLCLGIPPGECFG 1849

Qy 2086 LLGVNGAGKTSTFKMLTGDESTTGGEAFVNGHSLKELLQVQQSLGYCPQCDALFDELTA 2145  
 ||||| |||||:|:| | : ||| : ||:| :| :|:|||| |::|| ||  
 Db 1850 LLGVNGAGKTSTFRMVTGDTLPSSGEAVLAGHNVAQEPSAAHRSMGYCPQSDAIFDLLTG 1909

Qy 2146 REHLQLYTRLRGISWKDEARVVKWALE---KLELTKYADKPAGTYSGGNKRKLSTAIALI 2202  
 ||||:|: |||: ||:| : || :| | |||:||||| |||||:|:|:|:  
 Db 1910 REHLELFARLRGV---PEAQVAQTALSGVLRLGLPSYADRPAGTYSGGNKRKLATALALV 1966

Qy 2203 GYPAFIFLDEPTTGMDPKARRFLWNLIILDIKTGRSVVLTSHSMEECEALCTRLAIMVNG 2262  
 | || :||||| ||||| ||||| :| ::| ||||| ||||| ||||| |||||  
 Db 1967 GDPVVFLDEPTTGMDPSARRFLWNLLSVVREGRSVLTSHSMEECEALCTRLAIMVNG 2026

Qy 2263 RLRLGSIQHLKNRFGDGYMITVVRTKSSQSVKDVVRFFENRNFPEAMLKERHHTKVQYQL- 2321  
 | |||| ||||:||| | : :| | : : | ||:| |:| | ::::|  
 Db 2027 RFRCLGSAQHLKSRFGAGHTLTLRVPPDQP-EPAIAFIVTTFPDAELREVHGSRLRFQLP 2085

Qy 2322 KSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFVNFAKKQSDNLE--QQETE---- 2375  
 :||:| : :|:| |||||: ||: |:| | : | |||  
 Db 2086 PGGGCTLARVFRELAAQGAHGVEDFSVSQTTLEEVFLYFSKDQGEEEEGSGQETETREV 2145

Qy 2376 PPSALQSPLGCLLSLLRPRSAPT 2398  
 || | | | |  
 Db 2146 STPGLQHPKRVSRLFEDPSSVET 2168

RESULT 9

Q91V24

ID Q91V24 PRELIMINARY; PRT; 2159 AA.  
AC Q91V24;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE ATP-binding cassette transporter sub-family A member 7.  
GN ABCA7.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O1129, and DBA/2;  
RX MEDLINE=21328888; PubMed=11435699;  
RA Broccardo C., Osorio J., Luciani M.-F., Schriml L.M., Prades C.,  
RA Shulenin S., Arnould I., Naudin L., Lafargue C., Rosier M., Jordan B.,  
RA Mattei M.G., Dean M., Deneffe P., Chimini G.;  
RT "Comparative analysis of the promoter structure and genomic  
RT organization of the human and mouse ABCA7 gene encoding a novel ABCA  
RT transporter.";  
RL Cytogenet. Cell Genet. 92:264-270(2001).  
DR EMBL; AF287142; AAK56863.1; -.  
DR EMBL; AF287141; AAK56862.1; -.  
DR MGD; MGI:1351646; Abca7.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
DR GO; GO:0000166; F:nucleotide binding; IEA.  
DR GO; GO:0004601; F:peroxidase activity; IEA.  
DR GO; GO:0006979; P:response to oxidative stress; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR InterPro; IPR002016; Peroxidase.  
DR Pfam; PF00005; ABC\_tran; 2.  
DR ProDom; PD000006; ABC\_transporter; 2.  
DR SMART; SM00382; AAA; 2.  
DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
DR PROSITE; PS00435; PEROXIDASE\_1; 1.  
KW ATP-binding.  
SQ SEQUENCE 2159 AA; 236882 MW; CD2BE3FE0D8B822B CRC64;

Query Match 29.0%; Score 3675; DB 11; Length 2159;  
Best Local Similarity 35.4%; Pred. No. 1.9e-234;  
Matches 904; Conservative 371; Mismatches 733; Indels 544; Gaps 68;

Qy 6 QLQLLLWKNVTLKRRSPWVLA FEIFIP LVLFILLGLRQKKPTISVKEVPFYTAAPLTSA 65  
|| ||||| | :|| | | : || ||||| : :| | : | || ||  
Db 6 QLMLLLWKNYTYRRRQPIQLLVLLWPLFLFFILVAVRHSHPPLEHHECHF-PNKPLPSA 64  
Qy 66 GILPVMQSLCPDGQRDEF-----GFL-QYANSTVTQLLERLDRVVEEGNLFDPARP 115  
| :| :| | : | | | : :| : :|| | | | :



Db	65	GTVPWLQGLVCNVNNSCFQHPTPGKEKPGVLSNFKDSLISRLLAD-TRTVLGGHSIQDMLD	123
Qy	116	SLGSELEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSLPNS	175
Db	124	:   :    :     :       :    :	
Db	124	ALGKLIPVLR-----AVGGGARPGQESDQPT----SQGSVTKLLEKI-----	160
Qy	176	TAQALLAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELLLAPALL	235
Db	161	:   :   :   :   :      :	
Db	161	----LQRASLDP-----VLG-----QAQDSMRKFSDAIRDLA-----QELLTLPSLM	198
Qy	236	EQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRESGLSAELRNQLDVAK	295
Db	199	:      :   :	
Db	199	-----ELRALLRRPRGSAGSLELVSEALCS-----	223
Qy	296	VSQQLGLDAPNGSDSSPQAPPPrRLQALLGDLLDAQKVLQDQDVLVSALALLLPQGACTGR	355
Db	224	-----	223
Qy	356	TPGPPASGAGGAANGTGAGAV---MGPNATAEEGAPSAALATPD-TLQGQCSAFV----	407
Db	224	:     :   :              :	
Db	224	TKGPSSPG-GLSLNWEANQLNEFMGP-----EVAP-----ALPDNSLSPACSEFVGTLTD	272
Qy	408	-----QLWAGLQPILCGNRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKIL	461
Db	273	:	
Db	273	DHPVSRLLRRLKPLILG-----KIL	293
Qy	462	YAPAGSEVDRVILKANETFAFVGNVTHYAQVWLNISAEIRSFLQ-----	506
Db	294	:   : :   : :    : : :   : :	
Db	294	FAPDTNFTRKLMAQVNQTFEELALLRDLHELWGVLGPQIFNEMNDSTNVAMLQRLLDVGG	353
Qy	507	-GRLQQHLRWLQQYVAELRLHPEALNLSLDELPPALRQDNFSLPSGMALLQQLDTIDNAA	565
Db	354	:        : :      :     : :     : :	
Db	354	TGQRQQTPr-----AQKKL--EAIK---DFLDPS--RGYSWREAHADMGRLAGILG--	398
Qy	566	CGWIQFMSKVSVDIFKGFPEESIVNYTLNQAYQDNVTVFASVIF-----QTRKD	615
Db	399	:  :     ::  : : :   :   :	
Db	399	----QMMECVSLDKLEAVPSEEALVSRALELLGERR--LWAGIVFLSPEHPLDPSELSSP	452
Qy	616	GSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLY---GFVWIQDMMERAI	672
Db	453	:    : :    :       :     :     :    :    :	
Db	453	ALSPGHLRFKIRMDIDDVTRTNKIRDKFWDGPSPADPFMDLRYVWGGFVYLQDLLEQAAV	512
Qy	673	DTFVGHDVVEPGSYVQMPYPCYTRDDFLFVIEHMMPLCMVISWVYSVAMTIQHIVAEKE	732
Db	513	:             :    : :    :    :	
Db	513	RVLGGGN-SRTGLYLQMPHPCYVDDVFLRVLRSRLPLFLTLAWIYSVALTVKAVVREKE	571
Qy	733	HRLKEVMKTMGLNNAVHWVAFITGFVQLSISVTALTALIKYQVLMHSHVVIWFLAV	792
Db	572	:    :     :    :   : :     :     :   :     :	
Db	572	TRLRETMRAMGLSRAVLWLGLWFLSCLGPFLVSAALLVLVLKLGNIPLYPYSHPVVIFLFLAA	631
Qy	793	YAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPY-MYVAIREEVAHDKITAFEKCIA	851
Db	632	:   :   :  :   :    :    :   :      : :	
Db	632	FAVATVAQSFLLSAFFSRANLAAACGGLAYFALYLPVLCVAVWRERLHLGGLLA-----A	686
Qy	852	SLMSTTAFGLSKYFALYEAVGVIQWHTFSQSPVEGDDFNLLAVTMLMVDVAVYGILT	911
Db	687	:      :                 :   :    :	
Db	687	SLLSPVAFGFGCESLALLEEQDGAQWHNLGTGPAE-DVFSLAQVSAFLLLDAVIYGLAL	745

Qy 912 WYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSVMEEDQACAMES 971  
 ||:||| || ||:| || || ::||| | | | : ||: | | :  
 Db 746 WYLEAVCPGQYGIPEPWNFPFRRSYWCGPG-----PPKSSVL----APAPQD 788  
 Qy 972 RRFEETRGMEEEPTHLPLVVCVDKLTQVYKDDKKLALNKLNLNLYENQVVSFLGHNGAGK 1031  
 : || | | : | | :: : || | : | | : : |||||  
 Db 789 PKVL----VEEPPLGLVPGVSIRGLKKHFRGCPQPALQGLNLDYEGHITAFLGHNGAGK 844  
 Qy 1032 TTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSR 1091  
 |||:||||:||||:||||:| ||:|:| ||| |||||:| |||  
 Db 845 TTTLSILSGLFPSSGSASILGHDVQTNMAAIRPHLGICPQYNVLFDMLTVEEHVWFYGR 904  
 Qy 1092 LKSMAQEEIRREMDKMIEDLELSNKRHSVLQTLSSGGMKRKLSVAIAFVGGSRAILDEPT 1151  
 || :: : | ::| | : | : | : : ||||:||||| ||||| :|:| |||  
 Db 905 LKGVSAAMGPERERLIRDVGLTLKRDTQTRHLSGGMQRKLSVAIAFVGGSRVIMDEPT 964  
 Qy 1152 AGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPLFLKG 1211  
 ||||| :|| ||:|:| ||| :| ||||:| |||:| ||||:|: :: | | |||||:|  
 Db 965 AGVDPASRRGIWELLKYREGRTLILSTHHLDEAELLGDRVAMVAGGSLCCGSPLFLRR 1024  
 Qy 1212 TYGDGYRLTLVKRPAE-----PGGPQEP-----GLASS-----PP 1241  
 | || |||| | ::| | | | |  
 Db 1025 HLGCGYYLTLVKSSQSLVTHDAKGDSDEPRREKKSDGNGRTSDTAFTRGTSKSNQAPAP 1084  
 Qy 1242 GRAPLSSCSELQVSQFIRKHVASCLLVSDTSTELSYILPSEAACKGAFERLFQHLERSLD 1301  
 | |:: | :: : ::|| || | || :|| | | :| :|| |:: |:  
 Db 1085 GAVPITP-STARILELVQQHVPGAQLVEDLPHELLLVLPYAGALDGSFAMVFQELDQQLE 1143  
 Qy 1302 ALHLSSFGLMDTTL EEVFLKVSEEDQSLENSEADVKESRKDVLPGAEGPASGEGHAGNLA 1361  
 | | : | : || |||:|||| | : | ||  
 Db 1144 LLGLTGYGISDTNLEEIFLKVVED-----AHREG----- 1172  
 Qy 1362 RCSELTQSQASLQSASSVGSARGDEGAGYTDVYGDYRPLFD----NPQDPDNVSLQEVEA 1417  
 || || || | :|  
 Db 1173 -----GDSRQQLHLRTCTPQPPTGPEASVLEN 1199  
 Qy 1418 EALS RVG--QGSRKLDGGWLKVRQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTV 1475  
 | : | : :: | | :| || ||| |||: : ||:|:| || | :|  
 Db 1200 GELAPQGLAPNAAQVQGWTLTCQQLRALLHKRFLARRSRRLFAQVVLPAFLVGLALFF 1259  
 Qy 1476 ALSVPEIGDLPLVLSPSQYHNYTQPRGNFI PYANEERREYRLRLSP-DASPQQLVSTFR 1534  
 :| || | ||| |||: | | : : :| : | | : |:  
 Db 1260 SLIVPPFGQYPPQLSPAMY-----GPQVSFFSED-----APGDPNRMKLLLEALL 1304  
 Qy 1535 LPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPP 1594  
 :|: : | | | : || | | | :| | |  
 Db 1305 GEAGLQEPSMQDKDARG-----SECTHSLACYFTVPEVPPDVASILASGNWTPESP 1355  
 Qy 1595 SPAPSDSPASPEDDLQAWNVS LPPTAGPEMWT SAPSLPRLVREPVRCTCSAQGTGF---S 1651  
 ||| | || |  
 Db 1356 SPA-----CQCSQPGARRLLPD 1372  
 Qy 1652 CPSSVGGHPPQMRVV-TGDILTDITGHNVSEYLLFTSDRFR LH-----RYGAIT 1699  
 ||: || || | |:: : || |||:| : | ||| :  
 Db 1373 CPAGAGGPPPPQAVAGLGEVVQNL TGRNVSDFLVKTYPSLVRRGLKTKKWDEVRYGGFS 1432

Qy 1700 FGNVLKSIPASFGTRAPPMVRKIA-----VRRAAQV 1730  
 | :| :|| :| | : :  
 Db 1433 LGGRDPLPTGH-----EVVRTLAEIRALLSPQPGNALDRIINNLTQWALGLDARNSLKI 1487

Qy 1731 FYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPA--AYGITVTNHMPMNKTSASLSLDYLL 1788  
 ::|||:|:| :| || :| | | | :| || |||:| | || | :  
 Db 1488 WFNNKGWHAMVAFVNRANGLLHALLP---SGPVRHAHSITTLNHPLNLTKQLSEATLI 1544

Qy 1789 -QGTDVVIAIFIIVAMSFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWMDLN 1847  
 ||:::| :| ||||| :| :| :| :| ||| || :|| |::||| |  
 Db 1545 ASSVDVLVSICVVFAMSFVPASFTLVLEERITRAKHLQLVSGLPQTLYWLGNFLWDMCN 1604

Qy 1848 YLVPATCCVILFV-FDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWFEVPSSAYV 1906  
 ||| | | : :| :| || :| | ||||| :||| :| ||| :||  
 Db 1605 YLV-AVCIVVFIFLAFQQRAYVAPENLPALLLLLLLYGWSITPLMYPASFFFSVPSTAYV 1663

Qy 1907 FLIVINLFIGITATVATFLLQLFEHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEY 1966  
 | ||||| :::||:|:| |:::| :| || ||||:| |||:| | :  
 Db 1664 VITCINLFIGINSSMATFVLELLS-DQNLEQVSRILKQVFLIFPHFCLGRGLIDMVRNQA 1722

Qy 1967 INEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQMPVSTK 2026  
 : : : :| : :|| ||| : :|:| :| :| :| :| :| || |  
 Db 1723 MADAFERLGD-QQFQSPLRWDIIGKNLLAMMAQGFLITLLQHRNRLLPQSKPRLLP 1781

Qy 2027 PV-EDDVDVASERQRVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFG 2085  
 |: |:| ||| ||:|:| |:: : :|||: : | |||||: |||||  
 Db 1782 PLGEEDVDAQERERVTKGATQGDVLVLRDLTKVYRGQ---RNPVAVDRLCLGIPPGEFCG 1838

Qy 2086 LLGVNGAGKTSTFKMLTGDESTTGGEAFVNGHSVLELLQVQQSLGYCPQCDALFDELTA 2145  
 |||||:|:|:|:|:| : ||| : ||:| :| :|:| ||| ||:| ||  
 Db 1839 LLGVNGAGKTSTFRMVTGDTLPSSGEAVLAGHNVAQERSAAHRSMGYCPQSDAIFDLLTG 1898

Qy 2146 REHLQLYTRLRGISWKDEARVVKWALE---KLELTKYADKPAGTYSGGNKRKLSTAIALI 2202  
 |||:|:| |||: ||:| : || :| | ||:|||||||:|:|:|:|  
 Db 1899 REHLELFARLRGV---PEAQVAQTALSGLVRLGLPSYADRPAGTYSGGNKRKLATALALV 1955

Qy 2203 GYPAFIFLDEPTTGMDPKARRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNG 2262  
 | || :||| ||||| ||||| :| ::| ||||| ||||| ||||| |||||  
 Db 1956 GDPAVVFLDEPTTGMDPSARRFLWNSLLSVREGRSVVLTSHSMEECEALCTRLAIMVNG 2015

Qy 2263 RLRLCLGSIQHLKNRFGDGYMITVRTKSSQSVKDVVRFFNRNFPPEAMLKERHHTKVQYQL- 2321  
 | |||| ||| ||| |: :|:| | : :| || | |:| | ::::|  
 Db 2016 RFRLCLGSSQHLKGRFGAGHTLTLRVPPDQP-EPAIAFIRITFPGAELREVHGSRLRFQLP 2074

Qy 2322 KSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFNFAKKQSDNLEQQETEPSPALQ 2381  
 :| :|| :| :||:| |||: ||: |:| | : | |  
 Db 2075 PGGRCTLTRVFRELAAQGRAHGVEDFSVSQTTLEEVFLYFSKDQGEESSESSRQEAEEEEV 2134

Qy 2382 SPLGCLLSLLRPRSAPTELRLVADEPEDLDT 2413  
 | | | : :| | :|  
 Db 2135 SKPG-----RQHPKRVSRLFED-PSSVET 2157

RESULT 10

Q9BZC4

ID Q9BZC4 PRELIMINARY; PRT; 2146 AA.

AC Q9BZC4;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE ABC transporter member 7.  
 GN ABCA7.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21328888; PubMed=11435699;  
 RA Broccardo C., Osorio J., Luciani M.-F., Schriml L.M., Prades C.,  
 RA Shulenin S., Arnould I., Naudin L., Lafargue C., Rosier M., Jordan B.,  
 RA Mattei M.G., Dean M., Deneffe P., Chimini G.;  
 RT "Comparative analysis of the promoter structure and genomic  
 RT organization of the human and mouse ABCA7 gene encoding a novel ABCA  
 RT transporter.";  
 RL Cytogenet. Cell Genet. 92:264-270(2001).  
 DR EMBL; AF328787; AAK00959.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 2.  
 DR ProDom; PD000006; ABC\_transporter; 2.  
 DR SMART; SM00382; AAA; 2.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
 KW ATP-binding.  
 SQ SEQUENCE 2146 AA; 234306 MW; 2391728D5AD97E75 CRC64;

Query Match 28.6%; Score 3620; DB 4; Length 2146;  
 Best Local Similarity 35.8%; Pred. No. 8.8e-231;  
 Matches 899; Conservative 363; Mismatches 772; Indels 478; Gaps 60;

QY 1 MGFLHQLQLLLWKNVTLKRSPWVLAFEIFIPLVLFILLGLRQKKPTISVKEVPFYTAA 60  
 | | || ||||| :|| | | : || ||||| :| | : |  
 Db 1 MAFWTQLMLLLWKNFMYRRRQPVQQLLVEILLWPLFLFFILVAVRHSHPPLEHHECHF-PNK 59  
  
 QY 61 PLTSAGILPVMQSLCPDGQR-----DEFGFLQYANSTVTQLLERLDRVVEEGLNFD 111  
 || ||| :| :| | : :| | | | :| | | |  
 Db 60 PLPSAGTVPWLQGLICNVNNTCFQLTGPGEPPGRLSNFNDLSVSRLLADARTVLGGASAH 119  
  
 QY 112 PARPSLGSELEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLS 171  
 || : || |||  
 Db 120 RTLAGLGKLIATLR-----AARSTAQ----- 140  
  
 QY 172 LPNSTAQALLAARVDPP--EVYHLLFGPSSALDSQS---GLHKGQEPWSRLGGNPLFRME 226  
 | | |: | :|| :| || :| | :| | :| | |  
 Db 141 -PQPTKQSPL----EPPMLDVAELL---TSLLRTESLGLALGQAQEPHLSL----- 183  
  
 QY 227 ELLAPALLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARFSGLSAE 286  
 | | | :| || :| | | | :| :| | | :| :|

Db	184 -LEAAGDLAQELLALRSLVELRALLQRPGRGTSGLLELLSEALCS-----VRGPSST	233
Qy	287 LRNQLDVAKVSQQGLDAPNGSDSSPQAPPPrRLQALLGDLLDAQKVLQDQVDVLSALALL	346
	:   :   :   :   :   :   :   :   :   :   :   :	
Db	234 VGPSLNWYEASDLMELVGQEPEsalPDSSLSPACSELIGAL-----DSHPLSRL---	282
Qy	347 LPQGACTGRTPGPPASGAGGAANGTGAGAVMGPnATAEEGAPSAAALATPDTLQGQCsaF	406
Db	283 -----	282
Qy	407 VQLWAGLQPILCGNNRTIEPealRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAG	466
	:   :   :   :   :   :   :   :   :   :	
Db	283 --LWRRLKPLILG-----KLLFAPDT	301
Qy	467 SEVDRVILKANETFAfVGNVTHYAqVWLNIseIRSfLEqGRLqQHlRWLqqYVAELRLH	526
	: : :     : : :   :   :   :   :   :   :	
Db	302 PFTRKLMAQVNRTFEELtLLRDvREVwEMLGPRIftFMNDSSNVAMlQRllQMqDEGRrQ	361
Qy	527 P-----EALNLSDELPPAlRQDNfSLPSGMALLQQLDtIDNAACGWIQfMSK-VSV	577
	: : : :   :	
Db	362 PRPGGRDHMEALRSFLDP-----GSGGYSWQDAHADVGHLVGTLGRVTECLSL	409
Qy	578 DIFKGFPDEESIVNYTLNqAYQDNVTVFASVIfQTRKDGSLP-----HVHYKIR	627
	:     :   :   :   :   :   :   :   :   :	
Db	410 DKLEAAPSEAALVSRALQLLAEHR--FWAGVVFLGPEdSSDPTehPTDLGPGHVRIKIR	467
Qy	628 QNSSfTEKTNEIRRAYWRPGPN----TGGRFYfLYGFVWIQDMMERAIIdTFVGHdVVEP	683
	: :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	468 MDIDVVTRTNKIRDrfWDpgPAADPlTDLR-YVWGgFVYLQDLVERAAVRVLsgAN-PRA	525
Qy	684 GSYVQMFPYPcyTRDDfLVIEHmPLCMVISwVYSvAMTIQHIVAeKEHRLkeVMkTMG	743
	:                 : :   : : :   :   :   :   :   :	
Db	526 GLYLQQMPYPcyVDDVfLRLVLSRslPLfLTlaWiysvtLTVKAvVREKETRLrdTMRamG	585
Qy	744 LNNAVHWVAWFITGFVQLSiSVTaLTailKYqVLMHSHVVIIWLFLaVyAVAtIMFCFL	803
	:     :   : : :   :   :   :   :   :   :   :   :   :	
Db	586 LSRAVLWLGWfLSCLGPfLLSAAllVLVLKLGDIlPyShPGVVfLflAAFAVaTvTQSfL	645
Qy	804 VSVLYSKAKLASACGGIiyFLSyVPY-MYVAIREEvAHDKItAFekCIAsLMStTAfGLG	862
	:   : :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	646 LSAFFSRANLAACGGLayfSLyLPyVLCvaWr-----DRlpAGGrVAASLLSPVafGFG	700
Qy	863 SKYFaLYEVAGvgIQWhTfSQSpVEGDdFNLLLAVTmlMVdaVVYGilTWyIEavHPgmY	922
	:                       :   :   :   :   :   :   :   :	
Db	701 CESLALLEEQGEgaQwhNVGTTrPT-AdvfSLaqVSGLLLLDaLyGLatWyLeavCPgQY	759
Qy	923 GLPRPWYfPLQKSyWLGSgrTeAwEWSWPWARTPrLSVmEEdQAcameSrrFeetRGmEE	982
	:       :   :       :   :   :   :   :   :   :   :	
Db	760 GIPEPWNfPFRRsyWCgp-RppKSPaPCPtPlDPKvlV-----EE	798
Qy	983 EPThLPLVVCvdKLTKvyKddKKLalnKLSlnLYenQVVsFlghNgagKTTtmsIlTglF	1042
	: :   :   :   :   :   :   :   :   :   :	
Db	799 APPGLSPGVsvRSlekrfPgspQPaLRGLSlDFyQGHItaFLghNgagKTTtLSilSglF	858
Qy	1043 PPTSGSaTiYGHDirTEMdeIRKnLGMcPQHnVLFDrLTveEHlwFYsRLksMaqeeIRR	1102
	:         :   :   :   :   :   :   :   :   :   :   :	
Db	859 PPsgGSaFiLGHdvRSSmaAIRPhlgVcpQYNVLFdmLTvDeHvwfyGrLKGLSaaVvgP	918

Qy	1103	EMDKMIEDLELSNKRHSLVQTLSGGMRKRLSVAIAFVGGSRAIILDEPTAGVDPYARRAI	1162
Db	919	EQDRLLQDVGLVSKQSVQTRHLSGGMQRKLSVAIAFVGGSQVVLDEPTAGVDPASRRGI	978
Qy	1163	WDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPLFLKGTYG DG YRLTLV	1222
Db	979	WELLKYREGRTLILSTHHLDEAELLGDRVAVVAGGRLCCCGSPLFLRRHLGSGYYLTLV	1038
Qy	1223	KR--PAEPGGPQEPGLASSPPGRAPLSSCSE-----LQVSQFIRKHVASCLLVSDTSTE	1274
Db	1039	KARLPLTTNEKADTDMEGSVDTQRQEKNGSQGSRVGTPLQALLVQHWVPGARLVEELPHE	1098
Qy	1275	LSYILPSEAAKKGAFERLFQHLERSLDALHLSSFGIMDTTLEEVFLKVSEEDQSLENSEA	1334
Db	1099	LVLVLPYTGADHGSFATLFRDLRLAELRLTG YGISDTSLEEIFLKVVEE-----CAA	1152
Qy	1335	DVKESRKDVLPGAEGPASGEGHAG-NLARCELTQSQASLQSASSVGSAR-GEDEGAGYTD	1392
Db	1153	DT-----DMEDGSCGQHLCTGIAGLDVTLRLKMPPEALENGEPAGSAPETDQSGS---	1204
Qy	1393	VYGDYRPLFDNPQDPDNVSLQVEAEALSRVGQGSRKLDGGWLKVRQFHGLLVKRFHCAR	1452
Db	1205	-----PDAVG--RVQGWALTR-----QQLQALLKRFLLAR	1233
Qy	1453	RNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYHNYTQPRGNFIPIYANEE	1512
Db	1234	RSRRGLFAQIVLPALFVGLALVFSLIVPPFGHYPALRLSPTMY-----GAQVSFFSED	1286
Qy	1513	RREYRLRLSP-DASPPQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARF	1571
Db	1287	-----APGDPGRARLLEALLQEAG-----LEEP-----PVQHSSH-----RF	1318
Qy	1572	FDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDDELQAWNVSPLPPTAGPEMWTSAPSL	1631
Db	1319	SAPEVPAEVAKVLASGNWTPESPSPA-----	1344
Qy	1632	PRLVREPVRCTCSAQGTGF---SCPSSVGGHPPQMRVV-TGDILTITGHNVS EYLLFTS	1687
Db	1345	-----CQCSRPGARRLLPDCPAAAGGPPPPQAVTGSGEVVQNL TGRNLSDFLVKTY	1395
Qy	1688	DRFRLH-----RYGAITFG-----NVL	1704
Db	1396	PRLVRQGLKTKKWVNEVRYGGFSLGGRDPGLPSGQELGRSVEELWALLSPLPGGALDRVL	1455
Qy	1705	KSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPA	1764
Db	1456	KNLTA-----WAHSLDAQDSLKIWFNNKGWHSMAFVNRA SNAILRAHLPP---GPA	1504
Qy	1765	--AYGITVTNHPMNKTSASLSLDYLL-QGTDVVIAIFIIVAMSFVPASFVFLVAEKSTK	1821
Db	1505	RHAHSITTLNHPLNLTKEQLSEALMASSVDVLVSICVVFAMSFVPASFVFLV LIEERVTR	1564
Qy	1822	AKHLQFVSGCNPIIYWLANYVWDMNLNLYVPATCCVIIIFVFDLPAYTSPTNFPVLSLFL	1881
Db	1565	AKHLQLMGGLSPTLYWLG NFLWDMCNLYVPACIVVLI FLAFOORAYVAPANLPALLLLL	1624

Qy 1882 LYGWSITPIMYPASFWFEVPSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSY 1941  
 |||||:|||||:| |||:| | ||||| :|||:| | | : | :  
 Db 1625 LYGWSITPLMPASFFFSVPSTAYVVLTCINLFIGINGSMATFVLELFS-DQKLQEVSR 1683

Qy 1942 LKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKS PFWDIVTRGLVAMAVEGV 2001  
 || |||||: || ||:| | : : : : : | : || |:: | :|| :|  
 Db 1684 LKQVFLIFPHFCLGRGLIDMVRNQAMADAFERLGD-RQFQSPLRWEVVGKNLLAMVIQGP 1742

Qy 2002 VGFLLTIMCQYNFLRRPQRM---VSTKPV--EDDVDVASERQVRVLRGDADNDMVKIENL 2056  
 : | |:: | : | | : | | : | | | |::| | | : : ||  
 Db 1743 LFLFLTLLQH----RSQLLPQPRVRS LPLLGEEDVARERERVVQGATQGDVVLVRNL 1798

Qy 2057 TKVYKSRKIGRILAVDRCLGVRPGECFGLLG VNGAGKTSTFKMLTGDESTTGGEAFVNG 2116  
 ||||: : | : |||||: ||||| |||||:|:| | : || : |  
 Db 1799 TKVYRGQ---RMPAVDRCLGIPPGECFGLLG VNGAGKTSTFRMTGDTLASRGEAVLAG 1855

Qy 2117 HSVLKELLQVQOSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLEL 2176  
 ||| :| | :||| | | : | | |||: | : | : | |  
 Db 1856 HSVAREPSAAHLSMGYCPQSDAIFELLTGREHLELLARLGRVPEAQVATAGSGLARLGL 1915

Qy 2177 TKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLIIDLIKTG 2236  
 : |||:|||||:|:|:| | | :||||| | ||||| :| : : |  
 Db 1916 SWYADRPAGTYSGGNKRKLATALVGDPAVVFLDEPTTGMDPSARRFLWNSLLAVVREG 1975

Qy 2237 RSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVVRTKSSQSVKDV 2296  
 |||:|||||:|:|:| | | | | | | | | | | : : | : :  
 Db 1976 RSVMLTSHSMEECEALCSRLAIMVNGRFRCLGSPQHLKGRFAAGHTLTLRVPAARS-QPA 2034

Qy 2297 VRFFNRNFPPEAMLKERHHTKVQYQL-KSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLD 2355  
 | || : | : | : : : | : | : | : | : | : | : | :  
 Db 2035 AAFVAAEFPGSELREAHGGRLRFQLPPGGRCALARVFGELAVHGAEHGVEDFSVSQTMLE 2094

Qy 2356 NVEFVNFAK---KQSDNLEQQE----TEPPSALQSPLGCLLSLLRPRSAPTEL 2400  
 ||: | : | | | | : | | | | | : | : | |  
 Db 2095 EVFLYFSKDQKDEDETEEQKEAGVGVD PAPGLQHPRVVSQFLDDPSTAETVL 2146

RESULT 11

Q8IZY2

ID Q8IZY2 PRELIMINARY; PRT; 2146 AA.  
 AC Q8IZY2;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE ABC transporter ABCA7.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20549028; PubMed=11095984;  
 RA Kaminski W.E., Piehler A., Schmitz G.;  
 RT "Genomic organization of the human cholesterol-responsive ABC  
 RT transporter ABCA7: tandem linkage with the minor histocompatibility  
 RT antigen HA-1 gene.";  
 RL Biochem. Biophys. Res. Commun. 278:782-789(2000).

DR EMBL; AF311102; AAN04657.1; -.  
 DR EMBL; AF311057; AAN04657.1; JOINED.  
 DR EMBL; AF311058; AAN04657.1; JOINED.  
 DR EMBL; AF311059; AAN04657.1; JOINED.  
 DR EMBL; AF311060; AAN04657.1; JOINED.  
 DR EMBL; AF311061; AAN04657.1; JOINED.  
 DR EMBL; AF311062; AAN04657.1; JOINED.  
 DR EMBL; AF311063; AAN04657.1; JOINED.  
 DR EMBL; AF311064; AAN04657.1; JOINED.  
 DR EMBL; AF311065; AAN04657.1; JOINED.  
 DR EMBL; AF311066; AAN04657.1; JOINED.  
 DR EMBL; AF311067; AAN04657.1; JOINED.  
 DR EMBL; AF311068; AAN04657.1; JOINED.  
 DR EMBL; AF311069; AAN04657.1; JOINED.  
 DR EMBL; AF311070; AAN04657.1; JOINED.  
 DR EMBL; AF311071; AAN04657.1; JOINED.  
 DR EMBL; AF311072; AAN04657.1; JOINED.  
 DR EMBL; AF311073; AAN04657.1; JOINED.  
 DR EMBL; AF311074; AAN04657.1; JOINED.  
 DR EMBL; AF311075; AAN04657.1; JOINED.  
 DR EMBL; AF311076; AAN04657.1; JOINED.  
 DR EMBL; AF311077; AAN04657.1; JOINED.  
 DR EMBL; AF311078; AAN04657.1; JOINED.  
 DR EMBL; AF311079; AAN04657.1; JOINED.  
 DR EMBL; AF311080; AAN04657.1; JOINED.  
 DR EMBL; AF311081; AAN04657.1; JOINED.  
 DR EMBL; AF311082; AAN04657.1; JOINED.  
 DR EMBL; AF311083; AAN04657.1; JOINED.  
 DR EMBL; AF311084; AAN04657.1; JOINED.  
 DR EMBL; AF311085; AAN04657.1; JOINED.  
 DR EMBL; AF311086; AAN04657.1; JOINED.  
 DR EMBL; AF311087; AAN04657.1; JOINED.  
 DR EMBL; AF311088; AAN04657.1; JOINED.  
 DR EMBL; AF311089; AAN04657.1; JOINED.  
 DR EMBL; AF311090; AAN04657.1; JOINED.  
 DR EMBL; AF311091; AAN04657.1; JOINED.  
 DR EMBL; AF311092; AAN04657.1; JOINED.  
 DR EMBL; AF311093; AAN04657.1; JOINED.  
 DR EMBL; AF311094; AAN04657.1; JOINED.  
 DR EMBL; AF311095; AAN04657.1; JOINED.  
 DR EMBL; AF311096; AAN04657.1; JOINED.  
 DR EMBL; AF311097; AAN04657.1; JOINED.  
 DR EMBL; AF311098; AAN04657.1; JOINED.  
 DR EMBL; AF311099; AAN04657.1; JOINED.  
 DR EMBL; AF311100; AAN04657.1; JOINED.  
 DR EMBL; AF311101; AAN04657.1; JOINED.  
 DR Genew; HGNC:37; ABCA7.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 2.  
 DR ProDom; PD000006; ABC\_transporter; 2.  
 DR SMART; SM00382; AAA; 2.



DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
SQ SEQUENCE 2146 AA; 234422 MW; 33A128082D7B5BAF CRC64;

Query Match 28.6%; Score 3618; DB 4; Length 2146;  
Best Local Similarity 35.7%; Pred. No. 1.2e-230;  
Matches 898; Conservative 362; Mismatches 774; Indels 478; Gaps 60;

```
Qy      1 MGFLHQQLLLWKNVTLKRRSPWVLAFEIFIPVLVFFILLGLRQKKPTISVKEVPFYTAA 60
      | | | | | | | | | | : | | | | | : | | | | : | |
Db      1 MAFWTQLMLLLWKNFMYRRRQPVQLLVELLWPLFLFFILVAVRHSHPLEHHECHF-PNK 59

Qy     61 PLTSAGILPVMQSLCPDGQR-----DEFGFLQYANSTVTQLLERLDRVVEEGNLF 111
      | | | | : | : | : | : | | | | : | | | |
Db     60 PLPSAGTVPWLQGLICNVNNTCFPQLTPGEEPGRLSNFNDSLVSRLADARTVLGGASAH 119

Qy    112 PARPSLGSELEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLS 171
      | | : | | | | | | | | | |
Db    120 RTLAGLGKLIATLR-----AARSTAQ----- 140

Qy    172 LPNSTAQALLAARVDPP--EVYHLLFGPSSALDSQS---GLHKGQEPWSRLGGNPLFRME 226
      | | | | : | : | | | : | | : | | | | |
Db    141 -PQPTKQSPL----EPPMLDVAELL---TSLLRTESLGLALGQAQEPLHSL----- 183

Qy    227 ELLLAPALLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAE 286
      | | | | : | | | | | | | | : | | | | |
Db    184 -LEAAEDLAQELLALRSLVELRALLRPRGTSGPLELLSEALCS-----VRGPSST 233

Qy    287 LRNQLDVAKVSQQGLGLDAPNGSDSSPQAPPPRRQLALLGDLDDAQKVLQDQDVLALALL 346
      : | : | : | : | : | : | | | | | | | |
Db    234 VGPSLNWYEASDLMEVGQEPESALPDSSLSPACSELIGAL-----DSHPLSRL--- 282

Qy    347 LPQGACTGRTPGPPASGAGGAANGTGAGAVMGPNNATAEEGAPSAAALATPDTLQGQCSAF 406
Db    283 ----- 282

Qy    407 VQLWAGLQPILCGNNRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAG 466
      | | | | : | | | | | | | | | | | |
Db    283 --LWRRCLKPLILG-----KLLFAPDT 301

Qy    467 SEVDRVILKANETFAFVGNVTHYAQVWLNISAEIRSFLEQGRLQQHLRWLQQYVAELRLH 526
      : : : | | : : : | | : | : | : | | | |
Db    302 PFTRKLMAQVNRTFEELTLRDVREVWEMLGPRIFTFMNDSSNVAMLQRLQLMQDEGRRQ 361

Qy    527 P-----EALNLSLDELPPALRQDNFSLPSGMALLQQLDITIDNAACGWIQFMSK-VSV 577
      | | | | | | | | | | | | | | : : : :
Db    362 PRPGGRDHMEALRSFLDP-----GSGGYSWQDAHADVGLVGTGRVTECLSL 409

Qy    578 DIFKGFPDEESIVNYTLNQAYQDNVTVFASVIFQTRKDGSLPP-----HVHYKIR 627
      | : | | : | : | : | | : | | | | | | |
Db    410 DKLEAAPSEAAALVSRLQLLAEHR--FWAGVVFLGPELSSDPTEHPTPDLGPGHVRIKIR 467

Qy    628 QNSSFTEKTNEIRRAYWRPGPN---TGGRFYFLYGFVWIQDMMERAIIDTFVGHVVEP 683
      : : | | : | | | | | | | | | : | :
Db    468 MDIDVVTRTNKIRDREWDPGPAADPLTDLR-YVWGGFVYLQDLVERAAVRVLSGAN-PRA 525

Qy    684 GSYVQMFYPYPCYTRDDFLFVIEHMMPLCMVISWVYSVAMTIQHIVAEKEHRLKEVMKTMG 743
```

Db	526	GLYLQQMPYPCYVDDVFLRVLSRSLPLFLTLLAWIYSVTLTVKAVVREKETRLRDTMRAMG	585
Qy	744	LNNAVHWWAFITGFVQLSISVTALTAILKYGQVLMHSHVVI IWLFLAVYAVATIMFCFL	803
Db	586	LSRAVLWLGWFLSCLGPFLLSAALLVLVLKLGDLIPYSHPGVVFLFLAAFAVATVTQSFL	645
Qy	804	VSVLYSKAKLASACGGIIYFLSYVPY-MYVAIREEVAHDKITAFEKCIASLMSTTAFGLG	862
Db	646	LSAFFSRANLAAACGGLAYFSLYLPYVLCVAWR-----DRLPAGGRVAASLLSPVAFGFG	700
Qy	863	SKYFALYEAVAGVGIQWHTFSQSPVEGDDFNLLAVTMLMVDVAVYGILTWYIEAVHPGMY	922
Db	701	CESLALLEEQGEQAQWHNVGTRPT-ADVFSLAQVSGLLLLDAALYGLATWYLEAVCPGQY	759
Qy	923	GLPRPWYFPLQKSYWLGSGRTEAWESWPWARTPRLSVMEEDQACAMESRRFEETRGME	982
Db	760	GIPEPWNFPFRRSYWCGR-RPPKSPAPCPTPLDPKVLV-----EE	798
Qy	983	EPHLLPLVVCVDKLTKVYKDDKKLALNKLNLNLYENQVVSFLGHNGAGKTTTMSILTGLF	1042
Db	799	APPGLSPGVSVRSLKRFPGSPQPALRGLSLDFYQGHITAFGLGHNGAGKTTTSLISGLF	858
Qy	1043	PPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSRLKSMAQEEIRR	1102
Db	859	PPSGGSAFILGHDVRSMAAIRPHLGVCQYNVLFDMLTVDHVVWFYGRCLKGLSAAVVG	918
Qy	1103	EMDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAFVGGSRAIILDEPTAGVDPYARRAI	1162
Db	919	EQDRLLQDVGLVSKQSVQTRHLSGGMQRKLSVAIAFVGGSQVVILDEPTAGVDPASRRGI	978
Qy	1163	WDLILKYKPGRTILLSTHMHDEADLLGDRIAIISHGKLKCCGSPLFLKGTYG DG YRLTLV	1222
Db	979	WELLKYREGRTLILSTHHLDEAELLGDRVAVVAGGRGCCGSPLFLRRHLGSGYYLTLV	1038
Qy	1223	KR--PAEPGGPQEPGLASSPPGRAPLSSCSE-----LQVSQFIRKHVASCLLVSDTSTE	1274
Db	1039	KARLPLTTNEKADTDMEGSVDTROEKNGSQGSRVGTPQLLALVQHWVPGARLVEELPHE	1098
Qy	1275	LSYILPSEAARKGAFERLFQHLERSLDALHLSSFGLMDDTLEEVFLKVSEEDQSLENSEA	1334
Db	1099	LVLVLPYTGAHDGSFATLFRELDTRLAELRLTGYSIDTSLEEIFLKVVEE-----CAA	1152
Qy	1335	DVKESRKDVLPGAEGPASGEGHAG-NLARCSELTQSQASLQSASSVGSA-RGDEGAGYTD	1392
Db	1153	DT-----DMEDGSCGQHLCTGIAGLDVTLRLKMPPEQETALENGEPAGSAPETDQSGS---	1204
Qy	1393	VYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQFHGLLVKRFHCAR	1452
Db	1205	-----PDAVG--RVQGWA LTR-----QQQLQALLLKRFL LAR	1233
Qy	1453	RNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSYHNYTQPRGNFIPIYANEE	1512
Db	1234	RSRRGLFAQIVLPALFVGLALVFSLIVPPFGHPALRLSPTMY-----GAQVSFFSED	1286
Qy	1513	RREYRLRLSP-DASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSGSGESRLLAARF	1571

Db	1287	-----APGDPGRARLLEALLQEAG-----LEEP-----PVQHSSH-----RF	1318
Qy	1572	FDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDIDLQAWNVSPLPTAGPEMWTSPAPSL	1631
		:     :	
Db	1319	SAPEVPAEVAKVLASGNWTPESPSPA-----	1344
Qy	1632	PRLVREPVRCTCSAQGTGF---SCPSSVGGHPPQMRVV-TGDILTDTGHNVSLEYLLFTS	1687
		:         :   : : :     :   :   :	
Db	1345	-----CQCSRPGARRLLPDCPAAAGPPPPQAVTGSGEVVQNLGTGRNLSDFLVKTY	1395
Qy	1688	DRFRLH-----RYGAITFG-----NVL	1704
		:	
Db	1396	PRLVRQGLKTKKWVNEVRYGGFSLGGRDPGLPSGQELGRSVEELWALLSPLPGGALDRVL	1455
Qy	1705	KSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYSMTYLNLSNNAILRANLPKSKGNPA	1764
		: :   : : : : :       :   :       :	
Db	1456	KNLTA-----WAHSLDAQDSLKIWFNNKGWHSMAFVNRRASNAILRAHLPP---GPA	1504
Qy	1765	--AYGITVTNHPMNKTSASL-SLDYLLQGTDVVIAIFIIVAMSFVPASFVFLVAEKSTK	1821
		:         :     :   :     :   :       :   :   :	
Db	1505	RHAHSITTNLNHLNLTKEQLFEAALMASSVDVLVSI CVVFAMSFVPASFVFLVIEERVTR	1564
Qy	1822	AKHLQFVSGCNP I I YWLANYVWDMNLNLYLPATCCV I I L F V F D L P A Y T S P T N F P A V L S L F L	1881
		:   :   :       :             :	
Db	1565	AKHLQLMGGLSPTLYWLG N F L W D M C N Y L V P A C I V V L I F L A F Q Q R A Y V A P A N L P A L L L L L L	1624
Qy	1882	LYGWSITPIMYPASFWEFVPSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSY	1941
		:         :       :         :       :       :	
Db	1625	LYGWSITPLMYPASFFFSVPSTAYVVLTCINLFIGINGSMATFVLELFS-DQKLQEVSR I	1683
Qy	1942	LKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGV	2001
		: :       :     : : : : :   :     :   :   :	
Db	1684	LKQVFLIFPHFCLGRGLIDMVRNQAMADAFERLGD-RQFQSPLRWEVVGNLLAMVIQGP	1742
Qy	2002	VGFLTITMCQYNFLRRPQRM P---VSTKPV---EDDVDVASERQVRVLRGDADNDMVKIENL	2056
		:     : :     :     :         :     :	
Db	1743	LFLFLTLLQLH----RSQLLPQPRVRS L P L L G E E D E D V A R E R E R V Q G A T Q G D V L V L R N L	1798
Qy	2057	TKVYKSRKIGRILAVDRCLGVRPGECFGLLG V N G A G K T S T F K M L T G D E S T T G G E A F V N G	2116
		: :   :                               :     :	
Db	1799	TKVYRGQ---RMPAVDRCLGIPPGECFGLLG V N G A G K T S T F R M V T G D T L A S R G E A V L A G	1855
Qy	2117	HSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLEL	2176
		:   :             :                 :   :   :	
Db	1856	HSVAREPSAAHLSMGYCPQSDAIFELLTGREHLELLARLRGVPEAQVAQTAGSGLARLGL	1915
Qy	2177	TKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLI L D L I K T G	2236
		:     :                   :     :                   :   : :	
Db	1916	SWYADRPAGTYSGGNKRKLATALALVGDPVVFLDEPTTGMDPSARRFLWNLSLLAVVREG	1975
Qy	2237	RSVVLTS H S M E E C E A L C T R L A I M V N G R L R C L G S I Q H L K N R F G D G Y M I T V R T K S S Q S V K D V	2296
		:                 :                         : :   :	
Db	1976	RSVMLTSHSMEECEALCSRLAIMVNGRFRCLGSPQH L K G R F A A G H T L T L R V P A A R S - Q P A	2034
Qy	2297	V R F F N R N F P E A M L K E R H H T K V Q Y Q L - K S E H I S L A Q V F S K M E Q V S G V L G I E D Y S V S Q T T L D	2355
		:     : : :     :   :   :   :	
Db	2035	AAFVAAEFPGAELREAHGGRLRFQLPPGGRCALARVFGELAVHGAEHGVEDFSVSQTMLE	2094

Qy 2356 NVFVNFAK---KQSDNLEQQE----TEPPSALQSPLGCLLSLLRPRSAPTEL 2400  
 ||: |:| | | ||:| :| || | | :| | |  
 Db 2095 EVFLYFSKDQKGKDEDETEEQKEAGVGVDPAAGLQHPKRVSQFLDDPSTAETVL 2146

RESULT 12

Q9NR73

ID Q9NR73 PRELIMINARY; PRT; 2146 AA.  
 AC Q9NR73;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Macrophage ABC transporter.  
 GN ABCA7.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20334305; PubMed=10873640;  
 RA Kaminski W.E., Orso E., Diederich W., Klucken J., Drobnik W.,  
 RA Schmitz G.;  
 RT "Identification of a Novel Human Sterol-Sensitive ATP-Binding Cassette  
 RT Transporter (ABCA7).";  
 RL Biochem. Biophys. Res. Commun. 273:532-538(2000).  
 DR EMBL; AF250238; AAF85794.1; -.  
 DR GO; GO:0016021; C:integral to membrane; TAS.  
 DR GO; GO:0005524; F:ATP binding; TAS.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; TAS.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 2.  
 DR ProDom; PD000006; ABC\_transporter; 2.  
 DR SMART; SM00382; AAA; 2.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
 KW ATP-binding.  
 SQ SEQUENCE 2146 AA; 234469 MW; 679B16EB2D75FF0D CRC64;

Query Match 28.5%; Score 3616; DB 4; Length 2146;  
 Best Local Similarity 35.7%; Pred. No. 1.6e-230;  
 Matches 895; Conservative 364; Mismatches 777; Indels 474; Gaps 59;

Qy 1 MGFLHQLQLLLWKNVTLKRRSPWVLAFEIFIPVLVFFILLGLRQKKPTISVKEVPFYTAA 60  
 | | || ||||| :|| | | |: || |||||: :| | : | |  
 Db 1 MAFWTQLMLLLWKNFMYRRRQPVQLLVELLWPLFLFFILVAVRHSHPLEHHECHF-PNK 59  
 Qy 61 PLTSAGILPVMQSLCPDGQR-----DEFGFLQYANSTVTQLLERLDRVVEEGLNFD 111  
 || ||| :| :| | : :|| | | :| : | | |  
 Db 60 PLPSAGTVPWQLGLICNVNNTCFPQLTPGEEPGRLSNFNDLSVSRLLADARTVLGGASAH 119  
 Qy 112 PARPSLGSELEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLS 171  
 || : || |||  
 Db 120 RTLAGLGKLIATLR-----AARSTAQ----- 140

Qy 172 LPNSTAQALLAARVDPP--EVYHLLFGPSSALDSQS---GLHKGQEPWSRLGGNPLFRME 226  
| | | : | : | | : | | : | | : | | | |  
Db 141 -PQPTKQSPL----EPPMLDVAELL---TSLLRTESLGLALGQAQEPLHSL----- 183

Qy 227 ELLLPALLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAE 286  
| | | : | | : | | | : | : | | | : | :  
Db 184 -LEAAEDLAQELLALRSLVELRALLQRPRTSGPLELLSEALCS-----VRGPSST 233

Qy 287 LRNQLDVAKVSQQLGLDAPNGSDSSPQAPPPRRLQALLGDLDAQKVLQDQDVLVSALALL 346  
: | : : | : | : | : | : | | | | | | |  
Db 234 VGPSLNWYEASDLMELVGQEPESALPDSSLPACSELIGAL-----DSHPLSRL--- 282

Qy 347 LPQGACTGRTPGPPASGAGGAANGTGAGAVMGNATAEEGAPSAAALATPDTLQGQCSAF 406  
Db 283 ----- 282

Qy 407 VQLWAGLQPILCGNNRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAG 466  
| | | : | : | | : | : | | | : | : | |  
Db 283 --LWRRLKPLILG-----KLLFAPDT 301

Qy 467 SEVDRVILKANETFAGVGNVTHYAQVWLNISAEIRSFLEQGRLOQHRLWLQOYVAELRLH 526  
: : : : | | : : : | : | : | : | : | | | |  
Db 302 PFTRKLMAQVNRTFEELTLRDVREVWEMLGPRIFTMNDSSNVAMLQRLQLMQDEGRRQ 361

Qy 527 P-----EALNLSLDELPPALRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSK-VSV 577  
| | | | | | | | | | | | | | : : : : | :  
Db 362 PRPGGRDHMEALRSFLDP-----GSGGYSWQDAHADVGHLVGTGLGRVTECLSL 409

Qy 578 DIFKGFDPDEESIVNYTLNQAYQDNVTVFASVIFQTRKDGSLPP-----HVHYKIR 627  
| : | | : | : | : : | : | : | | | | | | |  
Db 410 DKLEAAPSEAALVSRAQLLAEHR--FWAGVVFLGPEDSSDPTEHPTDLGPGHVRIKIR 467

Qy 628 QNSSFTEKTNEIRRAYWRPGPN----TGGRFYFLYGFVWIQDMMERAIIDTFVGHVVEP 683  
: : : | : | : | | | | | | | : | : : | :  
Db 468 MDIDVVRTNKRIRDRFWDPGPAADPLTDLR-YVWGGFVYLQDLVERAAVRVLSPAN-PRA 525

Qy 684 GSYVQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSVAMTIQHIVAEKEHRLKEVMKTMG 743  
| | : | | | | | | | : : | : : : | | | | : | : |  
Db 526 GLYLQQMPYPCYVDDVFLRVLRSRSLPFLFTLAWIYSVTLTVKAVVREKETRLRDTMRAMG 585

Qy 744 LNNAVHWVAVFITGQVQLSISVTALTALIKYQVLMHSHVVIWLFLAVYAVATIMFCFL 803  
| : | | : | : : : | | : | | : | : | : : | | : | : | : | : |  
Db 586 LSRAVLWLGWFLSCLGPFLLSAALLVLVLKLGDLIPYSHPGVVFLFLAAFAVATVTQSFL 645

Qy 804 VSVLYSKAKLASACGGIIYFLSYVPY-MYVAIREEVAHDKITAFEKCIASLMSTTAFGLG 862  
: | : | : | | : | | | : | : | : | : | : | : | : |  
Db 646 LSAFFSRANLAAACGGLAYFSLYLPVLCVAWR-----DRLPAGGRVAASLLSPVAFGFG 700

Qy 863 SKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLAVTMLMVDVAVVYGILTWYIEAVHPGMY 922  
: | | | | | | | | | | : | : : | : | : | : | : |  
Db 701 CESLALLEEQGEQAQWHNVGTRPT-ADVFSLAQVSGLLLLDAALYGLATWYLEAVCPGQY 759

Qy 923 GLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSVMEEDQACAMESRRFEETRGME 982  
| : | | | : | | | | | | | : | : | : | : |  
Db 760 GIPEPWNFPPRRSYWCGP-RPPKSPAPCPTPLDPKVLV-----EE 798

Qy 983 EPTHLPLVVCVDKLTQVYKDDKKIALNKLNLNLYENQVVSFLGHNGAGKTTTMSILTGLF 1042

Db	799	APPGLSPGVSVRSLEKRFPGSPQPALRGLSLDFYQGHITAF LGHNGAGKTTLSILSGLF	858
Qy	1043	PPTS GSATIYGHDIRTEMDEIRK NLGMCPQHNVLFDR LTV E EHLWFYSRLKSMAQEEIRR	1102
Db	859	PPSGGSAFILGH D VRSSMAAIRPHLGVC PQYNVLF DMLTVDEH VWFYGR LKGLSAAVVG P	918
Qy	1103	EMDKMIEDLELSNKRHSLVQTL SGGMKRKLSVAIAFVGGSRAIILDEPTAGVDPYARRAI	1162
Db	919	EQDRLLQDVGLVSKQSVQTRHLSGGMQRKLSVAIAFVGGSQV VILDEPTAGVDPASRRGI	978
Qy	1163	WDLILKYKPGRTILLSTHMHDEADLLGDRIAIISHGKLKCCGSPLFLKGT YGDGYRLTLV	1222
Db	979	WELLLKYREGRTLILSTHHLDEAELLGDRVAVVAGGR LCCCGSPLFLRRHLGSGYYLTLV	1038
Qy	1223	KR--PAEPGGPQEPGLASSPPGRAPLSSCSE-----LQVSQFIRKHVASC LLVSDTSTE	1274
Db	1039	KARLPLTTNEKADTDMEGSV DTRQEKNGSQGSRVGT PQLLALVQHWP GARLVEELPHE	1098
Qy	1275	LSYILPSEA AKKGA FERLFQHLERSLDALHLS SFGLMDTTLEEVFLKVSEEDQSLENSEA	1334
Db	1099	LVLVLPYTGAHDGSFATL FREL DTRLAELRLTGYGISDTSLEE IFLKVVEE-----CAA	1152
Qy	1335	DVKESRKDVLPGAEGPASGE GHAG-NLARCSELTQSQASLQSASSVGSA-RGDEGAGYTD	1392
Db	1153	DT-----DMEDGSCGQHLCTGIAGLDVTLRLKMP PQTALENGEPAGSAPETDQSGS---	1204
Qy	1393	VYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRFHGLLVKRFH CAR	1452
Db	1205	-----PDAVG--RVQGWALTR-----QQLQALLKRFLLAR	1233
Qy	1453	RNSKALFSQILLPAFFVCVAMTVALSVPEIGDL PPLVLSPSQYHNYTQPRGNFI PYANEE	1512
Db	1234	RSRRGLFAQIVLPALFVGLALVFSLIVPPFGHYPALRLSPTMY-----GAQVSFFSED	1286
Qy	1513	RREYRLRLSP-DASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTINLSSGESRLLAARF	1571
Db	1287	-----APGDPGRARLLEALLQEAG-----LEEP-----PVQHSSH-----RF	1318
Qy	1572	FDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPD EDLQAWNVS L PPTAGPEMWT SAPSL	1631
Db	1319	SAPEVPAEVAKVLASGNWTPESPSPA-----	1344
Qy	1632	PRLVREPVRCTCSAQGTGF---SCPSSVGGHPPQMRVV-TGDILTDTIGHNVSEYLLFTS	1687
Db	1345	-----CQCSQPGARRLLPDCPAAAGGPPPPQAVTGSGEVVQNL TGRNLSDFLVKTY	1395
Qy	1688	DRFRLH-----RYGAITFG-----NVL	1704
Db	1396	PRLVRQGLKTKKWVNEVRYGGFSLGGRDPGLPSGQELGRSVEELWALLSPLPGGALDRV L	1455
Qy	1705	KSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSMP TYLNSLNNAILRANLPKSKGNPA	1764
Db	1456	KNLTA-----WAHSLDAQDSLKIWFNNKGWHS MVAFVN RASNAILRAHLPPGRAR-H	1506
Qy	1765	AYGITVTNHPMNKTSASL-SLDYLLQGT DVVIAIFIIVAMS FVPASFV VFLVAEKSTKAK	1823

Db 1507 AHSITTLNHPLNLTKEQLFEAALMASSVDVLVSICVVFAMSFVPASFTLVLIEERVTRAK 1566

Qy 1824 HLQFVSGCNPIIYWLANYVWDMLNLYLVPATCCVILFVFDLPAYTSPTNFPVLSLFLLY 1883  
 ||| : | : | : ||| | : ||| ||||| | : | | : | | : | | |||

Db 1567 HLQLMGGLSPTLYWLGNFLWDMCNLYLPACIVVLIFLAFQQRAYVAPANLPALLLLLLLLY 1626

Qy 1884 GWSITPIMYPASFWFEVPSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSYLK 1943  
 ||||| : ||||| : | ||| : ||| | ||||| : : ||| : : | | : | : | |

Db 1627 GWSITPLMYPASFFFSVPSTAYVVLTCINLFIGINGSMATFVLELFS-DQKLQEVSRILK 1685

Qy 1944 SCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVG 2003  
 ||||| : || ||| : | | : : : : : | : : || | : | : | : | : | :

Db 1686 QVFLIFPHFCLGRGLIDMVRNQAMADAFERLGD-RQFQSPLRWEVVGKNLLAMVIQGPLF 1744

Qy 2004 FLLTIMCQYNFLRRPQRM---VSTKPV--EDDVDASERQVRVLRGDADNDMVKIENLTK 2058  
 | | : | : | | : | | : | : | | | | : | : | | | | : | : | | | |

Db 1745 LLFTLLLQH---RSQLLPQPRVRSPLLGEEDEDVARERERVQGATQGDVVLVLRNLTK 1800

Qy 2059 VYKSRKIGRILAVDRLCLGVPRGECFGLLVNGAGKTSTFKMLTGDESTTGGEAFVNGHS 2118  
 || : : | : ||||| : ||||| ||||| ||||| : : ||| : ||| : |||

Db 1801 VYRGQ---RMPAVDRLCLGIPPGECFGLLVNGAGKTSTFRMVTGDTLASRGEAVLAGHS 1857

Qy 2119 VLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTK 2178  
 | : | | : ||||| ||| : | | ||||| | |||| : | : | : | | :

Db 1858 VAREPSAAHLSMGYCPQSDAIFELLTGREHLELLARLRGVPEAQVAQTAGSGLARLGLSW 1917

Qy 2179 YADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLIIDLIKTGRS 2238  
 ||| : ||||| ||||| : ||| : ||| | | : ||||| ||||| ||||| : | : : |||

Db 1918 YADRPAGTYSGGNKRKLATALALVGDPVVFLDEPTTGMDPSARRFLWNSLLAVVREGS 1977

Qy 2239 VVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVRTKSSQSVKDVVR 2298  
 | : ||||| ||||| : ||||| ||||| ||||| || | : : | : | : : | :

Db 1978 VMLTSHSMEECEALCSRLAIMVNGRFRCLGSPQHLKGRFAAGHTTLRVPAARS-QPAAA 2036

Qy 2299 FFNRFPEAMLKERHHTKVQYQL-KSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNV 2357  
 | | | : | : | | : : : | | : | : | : | : | : | : | : | : |

Db 2037 FVAAEFPGSELREAHGGRRLRFQLPPGGRCLARVFGELAVHGAEHGVEDFSVSQTMLEEV 2096

Qy 2358 FVNFAK---KQSDNLEQQE---TEPPSALQSPLGCLLSLLRPRSAPTEL 2400  
 | : | : | | | | : | : | : | | | | | | | | | | | | | | |

Db 2097 FLYFSKDQKDEDTEEQKEAGVGVDPAAGLQHPKRVSQFLDDPSTAETVL 2146

# RESULT 13

Q96S58

ID Q96S58 PRELIMINARY; PRT; 2008 AA.

AC Q96S58;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE ABCA-SSN.

GN ABCA7/ABCA-SSN.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE=21255283; PubMed=11355874;  
RA Tanaka A., Ikeda Y., Abe-Dohmae S., Arakawa R., Sadanami K.,  
RA Kidera A., Nakagawa S., Nagase T., Aoki R., Kioka N., Amachi T.,  
RA Yokoyama S., Ueda K.;  
RT "Human ABCA1 Contains a Large Amino-Terminal Extracellular Domain  
RT Homologous to an Epitope of Sjogren's Syndrome.";  
RL Biochem. Biophys. Res. Commun. 283:1019-1025(2001).  
DR EMBL; AB055390; BAB62294.1; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
DR GO; GO:0000166; F:nucleotide binding; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR Pfam; PF00005; ABC\_tran; 2.  
DR ProDom; PD000006; ABC\_transporter; 2.  
DR SMART; SM00382; AAA; 2.  
DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
KW ATP-binding.  
SQ SEQUENCE 2008 AA; 218617 MW; 226FF85C24230B90 CRC64;

Query Match 27.7%; Score 3515; DB 4; Length 2008;  
Best Local Similarity 37.3%; Pred. No. 7.5e-224;  
Matches 840; Conservative 351; Mismatches 699; Indels 360; Gaps 56;

Qy 268 VC--SGQAAARARRFSGLSAELRNQLDVAKVSQQLGLDAPNGSDSSPQAPPPRRLQALLG 325  
|| :||:| | :| | | | | | :||  
Db 2 VCLGTGQSA-----GPLVSVQNHCPPCGLSPQESLGLALGQAQEP-----LHSLLE 47

Qy 326 DLLD-AQKVLQDQDVLVSALALL-LPQGACTGRTPGPPASGAGGAANGTGAGAVMGPNNATA 383  
| ||::| :| || |::| | || : : | :||:  
Db 48 AAEDLAQELLALRSLVELRALLQRPRG-----TSGPLELLSEALCSVRGPSSTVGPSLNW 102

Qy 384 EEGAPSAAAL-----ATPD-TLQGQCSAFV-----QLWAGLQPILCGNNRTIEP 426  
| : : | || :| || : || |::| |  
Db 103 YEASDLMEVLVGQEPESALPDSSLPACSELIGALDSHPLSRLLRRLKPLILG----- 155

Qy 427 EALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETFAFVGNV 486  
|:|:| | : : : | || : :  
Db 156 -----KLLFAPDTPFTRKLMAQVNRTFEELTLL 183

Qy 487 THYAQVVLNISAIEIRSFLEQGRLQQHLRWLQQYVAELRLHP-----EALNLSLDEL 538  
:|| : |::| :| | | | | || ||  
Db 184 RDVREVWEMLGPRIFTFMNDSSNVAMLQRLQMQDEGRRQPRPGGRDHMEALRSFLDP-- 241

Qy 539 PALRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSK-VSVDIFKGFDPDEESIVNYTLNQA 597  
|| | | : : : :| : | | : :|  
Db 242 -----GSGGYSWQDAHADVGHLVGTGLGRVTECLSLDKLEAAPSEALVSRALQLL 291

Qy 598 YQDNVTVFASVIFQTRKDGSLPP-----HVHYKIRQNSSFTEKTNEIRRAYWRPG 647  
: :| |::| :| | | || ||| : :||:| :| ||  
Db 292 AEHR--FWAGVVFLGPEDSSDPTEHPTDLGPGHVRIKIRMDIDVVTRTNKIRDRFWDPG 349

Qy 648 PN----TGGRFYFLYGFVWIQDMMERAIIDTFVGHADVVEPGSYVQMFYPCYTRDDFLFV 703



Db	350	PAADPLTDLR-YVWGGFVYLQDLVERAAVRVLSGAN-PRAGLYLQQMPYPCYVDDVFLRV	407
Qy	704	IEHMMPLCMVISWVYSVAMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAVFITGFWQLSI	763
Db	408	LSRSLPLFLT LAWIYSVTLT VKAVVREKETRLRDTMRAMGLSRAVLWLWGLFSLCLGPFL	467
Qy	764	SVTALTAILKYGOVLMHSHVVIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYF	823
Db	468	SAALLVLVLKLG DILPYSHPGVVFLFLAAFAVATVTQSFLLSAFFSRANLAAACGG LAYF	527
Qy	824	LSYVPY-MYVAIREEVAHDKITAF EKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFS	882
Db	528	SLYLPYVLCVAWR-----DRLPAGGRVAASLLSPVAFGFGCESLALLEEQGEGAQWHNVG	582
Qy	883	QSPVEGDDFNLLAVTMLMVDVAVYGILT WYIEAVHPGMYGLPRPWYFPLQKSYWL GSGR	942
Db	583	TRPT-ADVFLAQVSGLLLLDAALYGLATWYLEAVCPGQYGIPEPWNFPFRRSYWC GP-R	640
Qy	943	TEAWESWPWARTPRLSVMEEDQACAMESRRFEETRGMEEEP THLPLVVCVDKLT KVYKD	1002
Db	641	PPKSPAPCPTPLDPKVLV-----EEAPPGLSPGVSVRSLEKRFP G	680
Qy	1003	DKKLALNKL SLNLYENQVVSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDE	1062
Db	681	SPQPALRGLSLDFYQGHITAF LGHNGAGKTTLSILSGLFPSPGGSATILGH DVRSSMAA	740
Qy	1063	IRKNLGMCPQHNVLFDRLTVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQ	1122
Db	741	IRPHLGVC PQYNVLFDM LTVDEHVWFYGR LKGLSAAVVGPEQDRLLQDVGLVSKQSVQTR	800
Qy	1123	TLSGGMKRKLSVAIAFVGGSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTH M	1182
Db	801	HLSGGMQRKLSVAIAFVGGSQVVILDEPTAGVDPASRRGIWELL LKYREGRTLILSTHHL	860
Qy	1183	DEADLLGDRIAIISHGKLKCCGSPFLKGT YGDGYRLTLVKR--PAEPGGPQEPGLASSP	1240
Db	861	DEAELLGDRVAVVAGGR LCCGSPFLRRHLGSGGYLTLVKARLPLTTNEKADTDMEGSV	920
Qy	1241	PGRAPLSSCSE-----LQVSQFIRKHVASCLLVSDTSTELSYILPSEA AKKGA FERLFQ	1294
Db	921	DTRQEKNGSQGSRVGT PQLLALVQHWVPGARLVEELPHELVLVLPYTGAHDG SFATLFR	980
Qy	1295	HLERSLDALHLSSFGLMDTTLEEVFLKVSEEDQSLENSEADV KESRKDVLPGAEGPASGE	1354
Db	981	ELDTRLAELRLTG YGISDTSLEEIFLKVVEE-----CAADT-----DMEDGSCGQHLCT	1029
Qy	1355	GHAG-NLARCSELTQSQASLQSASSVGSA-RGDEGAGYTDVYGDYRPLFDNPQDPDNVSL	1412
Db	1030	GIAGLDVTLRLKMPPQETALENGEPAGSAPETDQSGS-----PDAVG-	1071
Qy	1413	QEVEAEALSRVGQGSRKLDGGWLKV RQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVA	1472
Db	1072	-RVQGWALTR-----QQLOALLKRFLLARRSRGLFAQIVLPALFVGLA	1115
Qy	1473	MTVALSVPEIGDLPLVLSPSQYHNYTQPRGNFI PYANEERREYRLRLSP-DASPQQIVS	1531

Db 1116 LVFSLIVPPFGHYPALRLSPTMY-----GAQVSFFSED-----APGDPGRARLLE 1160

Qy 1532 TFRLP SGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVP 1591  
: | : | : | | : | : |

Db 1161 ALLQEAG-----LEEP-----PVQHSSH-----RFSAPVPAEVAKVLASGNWTP 1200

Qy 1592 PPPSPAPSDSPASPEDLQAWNVS LPPTAGPEMWT SAPSLPRLVREPVRCTCSAQGTGF- 1650  
| | | | | | |

Db 1201 ESPSPA-----CQCSQPGARRL 1217

Qy 1651 --SCPSSVGGHPPQMRVV-TGDILTITGHNVS EYLLFTSDRFR LH-----RYG 1696  
| | : : | | | : : : : | : : : : | | |

Db 1218 LPDCPAAAGGPPPPQAVTGSGEVVQNL TGRNLSDFLVKTY PRLVRQGLKTKKWVNEVRYG 1277

Qy 1697 AITFG-----NVLKSIPASFGTRAPPMVRKIAV 1724  
: | | | : | :

Db 1278 GFSLGGRDPGLPSGQELGRSVEELWALLSPLPGGALDRVLKNLTA-----WAHSLDA 1329

Qy 1725 RRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASL-S 1783  
: : : : : | | : | : | : | : | : | : |

Db 1330 QDSLKIWFNNKGWHSMAFVN RASNAILRAHLP PGRAR-HAHSITTLNHPLNLTKEQ LFE 1388

Qy 1784 LDYLLQGT DVVIAIFII VAMSFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVW 1843  
: | | : : | : | | | | : | : | : | : | : | : |

Db 1389 AALMASSVDVLVSICVVFAMSFVPASFTLV LIEERVTRAKHLQLMGGLSPTLYWLG NFW 1448

Qy 1844 DMLNYLVPATCCVII LFVFDLPAYTSPTNFP AVLSLFLLYGWSITPIMYPASFWEVPSS 1903  
| | | | | : | | : | | | : | | | | : | | | : |

Db 1449 DMCNYLVPACIVVLI FLAQQRAYVAPANLPALL LLLLYGWSITPLMYPASFFFSVPST 1508

Qy 1904 AYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSY LKSCFLIFPNYNLGHGLMEMAY 1963  
| | | | | : : | : | : | : | : | : | : | : |

Db 1509 AYVVLTCINLFIGINGSMATFVLELFS-DQKLQEVSRILKQVFLIFPHFCLGRGLIDMVR 1567

Qy 1964 NEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQ RMP- 2022  
| : : : : : : : | : : | : : | : : | : : | : : |

Db 1568 NQAMADAFERLGD-RQFQSPLRWEVVGKNLLAMVIQGPLFLLFTLLLQH----RSQ LLPQ 1622

Qy 2023 --VSTKPV--EDDVDVASERQVR LRGDADNDMVKIENLTKVYKSRKIGRILAVDR LCLGV 2078  
| : | : | | | | : : : | : : : | : : : | : : : |

Db 1623 PRVRS LPLGEEDEDVARERERVVQGATQGDVLVLRNLTKVYRGQ---RMPAVDR LCLGI 1679

Qy 2079 RPGECEG L LGVNGAGKTSTFKMLTGDESTTGGEAFVNGH SVLKELLQVQQSLGYCPQ CDA 2138  
| | | | | : : | : | : | : | : | : | : |

Db 1680 PPGECEG L LGVNGAGKTSTFRMTGDTLASRGEAVLAGHSVAREPSAAHLSMGYCPQSDA 1739

Qy 2139 LFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYS GGNKRKLSTA 2198  
: | : | | : | | : | : | : | : | : | : | : |

Db 1740 IFELLTGREHLELLARLRGVPEAQVAQTAGSGLARLGLSWYADRPAGTYS GGNKRKLATA 1799

Qy 2199 IALIGYPAFI FLDEPTTGMDPKARRFLWN LILD LIKTGRSVVLTSHSMEECEALCTRLAI 2258  
: | : | : | : | : | : | : | : | : | : | : |

Db 1800 LALVGDP AVVFLDEPTTGMDPSARRFLWN SLLAVVREGRSVMLTSHSMEECEALCSRLAI 1859

Qy 2259 MVNGRLRCLGSIQHLKNRFGDGYMITVRTKSSQSVKDVVRFFNRNFP EAMLKERHHTKVQ 2318  
| | | | | : | : | : | : : : : | : | : | : |

Db 1860 MVNGRFRCLGSPQHLKGRFAAGHTLTLRVPAARS-QPAAAFVAAEFPGSELREAHGGRLR 1918

```

Qy      2319 YQL-KSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFNFAK---KQSDNLEQQE- 2373
          :||      :||:| | :|      | :||:| | | | | :| :| :| | | | | :|
Db      1919 FQLPPGGRCALARVFGELAVHGAEHGVEDFSVSQTMLEEVFLYFSKDQGKDEDTEEQKEA 1978

Qy      2374 ---TEPPSALQSPGLCLLSLLRPRSAPEL 2400
          :|      || |      | | :| | |
Db      1979 GVGVDPA PGLOHPKRVSOFLDDPSTAETVL 2008

```

## RESULT 14

Q86UK0

```

ID      Q86UK0          PRELIMINARY;      PRT; 2595 AA.
AC      Q86UK0;
DT      01-JUN-2003 (TrEMBLrel. 24, Created)
DT      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      ABCA12 transporter subfamily A.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=22583451; PubMed=12697999;
RA      Annilo T., Shulemin S., Chen Z.Q., Arnould I., Prades C., Lemoine C.,
RA      Maintoux C., Devaud C., Dean M., Deneffe P., Rosier M.;
RT      "Identification and characterization of a novel ABCA subfamily member,
RT      ABCA12, located in the lamellar ichthyosis region on 2q34.";
RL      Cytogenet. Genome Res. 98:169-176(2002).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Annilo T., Shulenin S., Chen Z.Q., Arnould I., Prades C., Lemoine C.,
RA      Maintoux C., Devaud C., Dean M., Deneffe P., Rosier M.;
RL      Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; AY219711; AAP21093.1; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0005524; F:ATP binding; IEA.
DR      GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR      GO; GO:0000166; F:nucleotide binding; IEA.
DR      GO; GO:0006810; P:transport; IEA.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR003439; ABC_transporter.
DR      Pfam; PF00005; ABC_tran; 2.
DR      ProDom; PD000006; ABC_transporter; 2.
DR      SMART; SM00382; AAA; 2.
DR      PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR      PROSITE; PS50893; ABC_TRANSPORTER_2; 2.
SQ      SEQUENCE      2595 AA; 293148 MW; A771C73A4276A238 CRC64;

```

Query Match 22.4%; Score 2831.5; DB 4; Length 2595;  
Best Local Similarity 30.1%; Pred. No. 3.3e-178;  
Matches 740; Conservative 432; Mismatches 814; Indels 469; Gaps 66;

QY 93 VTQLL---ERLDRVVEEGNLFDPARPSLGSELEALRQHLEALSAGPGTSGSHLDRSTVSS 149  
:|:| | : : | | : | | | | | : | | | |  
Db 436 LTELLCESETFSLIEKSCQLSDMSFGSLCEESEFDLQLEAAELGTEIAASLLYHDNVIS 495

Qy 150 FSL-DSVARNPQELWRFLTQNLSPNSTAQALLAARVDPPEVYHLLFGPSSALDSQSGH 208  
 : | : : | :: ||:: ||| :: : | : : | :  
 Db 496 KKVRDLLTGDPSKI-----NLNMDQFLEQALQMNYLE--NITQLIPIIEAMLHVNSAD 547  
 Qy 209 KGQEPWSRLGGNPLFRMEELLLAPALLEQLTCTPGSGE--LGRILTVPESQKGA----- 260  
 ::| | :|: | | | | | : ::| :| |  
 Db 548 ASEKPGQLL---EMFKNVE-----ELKEDLRRTTGMSNRTIDKLLAIPIDNRAEIISQV 599  
 Qy 261 -----LQGYRDAVCSGQAAARARRESGLSAELRNQLDVAKVSQQLGLDAPN 306  
 | : | : | : | : | : | : | : | :  
 Db 600 FWLHSCDTNITTPKLEDAMKEFCNLSLSERSRQSYLIGLTLHYLNIYNFTDKVFFPRKD 659  
 Qy 307 GSDSSPQAPPPR-----RLQALLGD-----LLDAQVLQDQVDVLSALALL----- 346  
 | | : ||: :| ||| : | : : : : |  
 Db 660 -----QKPVEKMMELFIRLKEILNQMASGTHPLLDKMRSILKQMHLPERSVPLTQAMYRSN 713  
 Qy 347 ---LPQGACTGRTPGPPASGAGGAANGTGAGAVMGPNATAEEGAPS-----A 390  
 ||| : : : | | | | : : | : |  
 Db 714 RMNTPQGSFSTISQALCSQGI-----TTEYLTAMLPSQRPKGNHTKDFLTyKLTKEQIA 768  
 Qy 391 AALATPDTLQGCsafvQ-----LWAGLQPILCGNNRTIEPEALRRGNMSSLGFT 440  
 : | : : : || | : | | |  
 Db 769 SKYGIPINTTPFCFSlyKDIINMPAGPVIWAFKPMllG----- 807  
 Qy 441 SKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVWLNIS-AE 499  
 : ||: || : : | : | : : | : |  
 Db 808 -----RILHAPYNPVTKAIMEKSNVTlRQLAELREKSQEWMDKSPLF 849  
 Qy 500 IRSF--LEQG--RLQQHLR--WLQQYVA-ELRLHPEALNLSLDELPPALRQDNFSLPSGM 552  
 : || | | || || ::| :| : | | : ||| || | : :  
 Db 850 MNSFHLNQAIPMLQNTLRNPVFQVFKFSVGLDAVELLKQIDEL-DILR---LKLENNI 905  
 Qy 553 ALLQQLDTIDNAACGWIQFMSKVSVDIFKGFpDEESIVNYTLNQA-----Y 598  
 : : ||: | : | : : | | | | |  
 Db 906 DIIDQLNT-----LSSLTVNI-----SSCVLYDRIQAaKTIDEMEREAKRLY 947  
 Qy 599 QDNVTVFASVIFQTRKdGS-----LPPHVHYKIRQNSSFTEKTNEIRRAYWRPG 647  
 : | : | |||: : | ||| : | || : : | : | ||  
 Db 948 KSN-ELFGSVIFKLPSNRSWHRGYDSGNVFLPPVIKYTIRMSLKTAQTTRSLRtkIWAPG 1006  
 Qy 648 PNTGGRFYFLYG--FVWIQDMMERAIIDTFVGHdVVEPGSYVQMFYPYCYTRDDFLFVIE 705  
 | : : || | : : || : ||| : | : | || |||: : : || :  
 Db 1007 PHNSPSHNQIYGRAFIYLQDSIERAIIElQTGRNSQEIavQVQAIPYPCFMKDNFLTsvS 1066  
 Qy 706 HMMPLCMVISWVYSVAMTIQHIVAEKEHRLKEVMKTMGLNNAVHwVawFITGFVQLSISV 765  
 : :| : : : || : | : : | ||: || | |||: | : || | : :  
 Db 1067 YSLPIVLMVAwVVFIAAFVKKLVYEKDLRLHEYMkMMGVNSCSHFFAWLIESVGFLlVTI 1126  
 Qy 766 TALTAILKYGQVLmHSHVVIWFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLS 825  
 |. |||: | : : | : : : | : : || : : : | : | : :  
 Db 1127 VILIIILKFGNlLPKTNGFIlFLYFSdYSFSVIAMSYLISVFFNNTNIAALIGSLIYIIA 1186  
 Qy 826 YVPYMYVAIREEVaHDKITAFEkCIASlMSTTAfGLGskyFALYEVAGVGiQWHTFSQSP 885  
 : ||: : | : : | ||: || | | | | : | : || ||  
 Db 1187 FFPFIVLVtVE----NELSYVLKVFMSLLSPTAFsYASQYIARYEEQGiGLQWENMYTSP 1242

Qy	886	VEGDDFNLLAVTMLMVD	AVVYGILTWYIEAVH	PGMYGLPRPWYFPLQKSYW---	LGSGR	942
Db	1243	VQDDTT\$FGWLCCILADS	FIYFLIAWYVRNVFPGTYGMAAPWYFPI	LPSYWKERFGCAE		1302
Qy	943	TEAWESWPWARTPRLS-----	VMEEDQACAMESRRFEETR	GMEEETHLPLVVCVD		994
Db	1303	VK-----	PEKSNGLMFTNIMQNTNPSASPEYMF--	SSNIEPEPKDLTVGVALH		1349
Qy	995	KLTKVYKDDKKLALNKL	SLNLNYENQVVSFLGHNGAGKTTTMSILT	TGLFPPTSGSATIYGH		1054
Db	1350	GVTKIY--	GSKVAVDNLNLNFYEGHITSLLGPNGAGKTTT	ISMLTGLFGASACTIFVYGK		1407
Qy	1055	DIRTEMDEIRKNLGMCPQ	HNVLFDRLTVEEHLWFYSRLK--	SMAQEEIRREMDKMIEDLE		1112
Db	1408	DIKTDLHTVRKNMGVCMQ	HDVLF\$YLT\$TKEHLLLYG\$IKVPHWTKKQLH	EEVKRTLKDTG		1467
Qy	1113	LSNKRHSLVQTL\$GGMKR	K\$LSVAIAFVGGSRAIILDEPTAGVDPYARRAI	WDLILKYKPG		1172
Db	1468	LYSHRHKRVGTL\$GGMKR	KL\$ISIALIGGSRVVILDEPSTGVDPC\$RRSI	WDV\$ISKNTA		1527
Qy	1173	RTILLSTHHMDEADLLG	DRIAIISHGKLKCCGSPLFLKGTYGDGYRLTL	VKRPAEPGGPQ		1232
Db	1528	RTIILSTHHLDEAEVLS	DRIAFLEQGGRLCCGSPFYLKEAFGDGYHLTL	TKK-----K		1580
Qy	1233	EPGLASSPPGRAPLSSC	SELQVSQFIRKHVASCLLVSDTSTELSYILP	SEAAK-KGA	FER	1291
Db	1581	SPNLNAN-----	AVCDTMAVTAMIQSHLPEAYLKEDIGGELVY	VLPPFSTKVS	GAYLS	1633
Qy	1292	LFQHLERSLDALHLSS	FGLMDTTLEEVLKVSEEDQ\$LENSEADVKESR	KDVLPGAEGPA		1351
Db	1634	LLRALDNGMGDLNIG	CGYISD\$T\$VEEVFLNLTKESQ--	KNSAMSLE-----		1677
Qy	1352	SGEGHAGNLARCELTQ	SQASLQSASSVGSARGDEGAGYTDVYGDYR	PLFDNPQDPDNVS		1411
Db	1678	-----	HLTQKKIGNSNANGIST-----	-----	PDDL\$	1699
Qy	1412	LQ-----	EVEAEALSRVGQGSRKLDGGWLKVRQFHGLL	VKR\$HFCARRNSKALFSQILLPA		1466
Db	1700	VSSSNFTDRDDKILTR---	GERLDGFGLL\$K\$KIMAILIKRFHH\$TRRNW	KGLIAQVILPI		1755
Qy	1467	FFVCVAMTVALSVPEI	GDLPPLVLSP\$QYHNYTQPRGNFI	PYANEERREYRLRLSPDASP		1526
Db	1756	VFVTTAMGLGTLRNS	SN\$YPEIQISPSLYG--	TSEQTAF--	YANYH-----	PST 1800
Qy	1527	QQLVSTFRLPSGVGAT	CVL\$KSPANGSLGPTLNLSSGESRLLAAR	FFDSMCLESFTQGLPL		1586
Db	1801	EALVSAMWDFPGI-----	-----	DNMCLNT-----		1820
Qy	1587	SNFVPPPPSPAPSD	SPASPD\$DLQAWNVSIPPTAGPEMWT\$APSL	PRLVREPVR----	CT	1642
Db	1821	-----	SDLQCLNKDSLEKWNT\$-----	-----	GEPITNFGVCS	1848
Qy	1643	CSAQGTGFSCPSSV	GGHPPQMRVVTGDILTITGHNVSEYLLFT	SDR\$RLHRYGAITFGN		1702
Db	1849	CSENVQ--	ECP-KFNYSPPHRTYSSQVIYNLTGQR	VENYLISTANEFVQKRYGGWSFGL		1905
Qy	1703	VL-----	KSIPASFGTRAPPMVRKIAVRRAAQV	FYNNKGYSMPTYLN	SLNNAILRA	1754

		:  :		: : :	:       :		:		
Db	1906	PLTKDLRF	DITGVPAN-----	RTLAKVWYD	PEGYHSL	PAYLNSL	NNFLLRV 1951		
Qy	1755	NLPKSKGN	PAYGITVTNHP----	MNKT	SASLSL	DYLLQGT	DVVIAIFIIVAMSFVPASF 1810		
		:	:	:     :	:	:   :	:     :		
Db	1952	NM--	SKYDAARHGIIM	YSHYPY	PGVQDQEQ	ATIS-----	SLIDILVALSILMGYSVTTASF 2004		
Qy	1811	VVFLVAEK	STKAKHLQFV	SGCNPII	YWLANYV	WDMNLN	YLPATCCVILFVFDLPAYTSP 1870		
		:	:	:	:	:   :     :	:		
Db	2005	VTYV	REHQTKAKQLQ	HISGIG	VTCTYWVT	NFIYDM	VFYLVPAFVSIGIIAIFKLPAFYSE 2064		
Qy	1871	TNFP	PAVLSLFL	LYGWSIT	PIMYPAS	FWFEVP	SSAYVFLIVINLFIGITA----TVATFLL 1926		
				:	:	:	:		
Db	2065	NNLG	AVSL	LLLLL	FGHATF	SWMYLL	AGLFHETGMAFITYVCVNLFFGINSIVSLSVVYFLS 2124		
Qy	1927	QLFEHDK	DLKVNSYL	KSCFLI	FPNPNL	GHGLME	MAYNEYINEYYAKIGQFDMKSPFEW 1986		
		:	:	:	:	:       :	: : : :   :		
Db	2125	KEKP	NDPTLE	LISE	TLKRIF	LIFPQ	FCFGYGLIELSQQSVLDFLKAYG-VEYPNETFEM 2183		
Qy	1987	DIVTR	GLVAMAVE	GVVGFLL	TIMCQY	NFLR-----	RPQRMVSTKPVEDDDVDVASERQR 2040		
		:	:	:	:     :	: : :	: : :		
Db	2184	NKLG	AMFVAL	VSQGT	MFFSL	RLLINE	SLIKKLRLFFRKFNSSHVRETIDEDEDVRAERLR 2243		
Qy	2041	VLRG	DADNDM	VKIENL	TKVYK	SRKIGR	ILAVDRLCLGVRPGEFCGLLGVNGAGKTSTFKM 2100		
			:	:   : :	:	:   :     :	: :		
Db	2244	VESGA	AEDLVQ	LYCL	TKTYQ	LIH-KK	IIAVNNISIGIPAGECFGLLGVNGAGKTTFIKM 2302		
Qy	2101	LTGD-----	ESTTGGE	AFVNGH	SVLKELL	QVQQSL	GYCPQCDALFDELTAREHLQ 2150		
			:		:	:			
Db	2303	LTGD	IIPSSG	NILIRN	KTGSLG	HVDSSH	SSL-----VGYPQEDALDDLVTVEEHLV 2353		
Qy	2151	LYTRL	RGISWK	DEARV	VVKWAL	EKLELT	KYADKPAGTYSGGNKRKLSTAIALIGYPAFIFL 2210		
		:			:	:   :			
Db	2354	FYAR	VHGIPE	KDIKET	VHKLL	RRLHLM	PFKDRATSMCSYGTKRKLSTALALIGKPSILL 2413		
Qy	2211	DEPT	TGMDPK	ARRFLW	NLILD	LIKTGR	SVVLTSHSMEECEALCTRLAIMVNGRLRCLGSI 2270		
		:	:	:	: : :	:	: :       :		
Db	2414	DEP	SSGMDP	KSKRHL	WKIISE	VQNKCS	VILTSHSMEECEALCTRLAIMVNGKFQCIGSL 2473		
Qy	2271	QHLKN	RFGDGY	MITV	RTKSSQ	-SVKD	VVRFFNRNFP	EAMLERHHTKVQYQLKSEHISLA 2329	
		:	:	: : :	: :	:     :	:     :   :		
Db	2474	QH	IKSR	FGRG	FTVKV	HLKNNK	VTMETLT	KFMQLHFPKTYLKDQHL	SMLEYHVPVTTAGGVA 2533
Qy	2330	QVFS	KMEQV	SGVLG	IEDYSV	SQTTL	DNVFNFAKKQSDNLEQQETEP	PPSALQSPL 2384	
		:   :	:	:	:	:	:		
Db	2534	NIFD	LLETN	KTALN	ITN	FLVSQ	TTL	EEVF	INFADQ----KSYETADTSSQGSTI 2584

RESULT 15

Q8IZW6

ID Q8IZW6 PRELIMINARY; PRT; 2347 AA.  
AC Q8IZW6;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE ATP-binding cassette transporter family A member 12.  
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Schaap F.G., van Wijland M., Groen A.K.;  
 RT "Cloning of a novel ABC transporter (ABCA12) tentatively involved in  
 RT lipid homeostatis.";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF418105; AAN40735.1; -.  
 DR Genew; HGNC:14637; ABCA12.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 2.  
 DR ProDom; PD000006; ABC\_transporter; 2.  
 DR SMART; SM00382; AAA; 2.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
 KW ATP-binding.  
 SQ SEQUENCE 2347 AA; 264963 MW; 9B6E13FD0F0F67AD CRC64;

Query Match 22.3%; Score 2827.5; DB 4; Length 2347;  
 Best Local Similarity 30.2%; Pred. No. 5.1e-178;  
 Matches 741; Conservative 432; Mismatches 813; Indels 469; Gaps 66;

Qy 93 VTQLL---ERLDRVVEEGNLFDPARPSLGSELEALRQHLEALSAGPGTSGSHLDRSTVSS 149  
 :|:| | : : | | : | | | | | : | | | |  
 Db 188 LTELCESETFSLIEKSCQLSDMSFGSLCEESEFDLQLLEAAELGTEIAASLLYHDNVIS 247  
  
 Qy 150 FSL-DSVARNPQELWRFLTQNLSPNSTAQALLAARVDPPEVYHLLFGPSSALDSQSGLH 208  
 : | : : | : : | | : : | | : : : | : :  
 Db 248 KKVRDLLTGDPSKI-----NLNMDQFLEQALQMNYLE--NITQLIPIIEAMLHVNSAD 299  
  
 Qy 209 KGQEPWSRLGGNPLFRMEELLLAPALLEQLTCTPGSGE--LGRILTVPESQKGA----- 260  
 : : | | : : | | : | | | | | : : : | : |  
 Db 300 ASEKPGQLL---EMFKNVE-----ELKEDLRRTTGMSNRTIDKLLAIPDPNRAEIIISQV 351  
  
 Qy 261 -----LQGYRDAVCSGQAAARARRFSGLSAELRNQLDVAKVSQQGLGDAPN 306  
 | : | : : | : : | : | : : : : :  
 Db 352 FWLHSCDTNITTPKLEDAMKEFCNLSLSERSRQSYLIGLTLHLYLNIYNFTYKVFFPRKD 411  
  
 Qy 307 GSDSSPQAPPPR-----RLQALLGD-----LLDAQVLQDQVDVLSALALL----- 346  
 | | : | | : | | | | : | : : : : |  
 Db 412 -----QKPVEKMMELFIRLKEILNQMASGTHPLLDKMRSCLKQMHLPKSVPLTQAMYRSN 465  
  
 Qy 347 ---LPQGACTGRTPGPPASGAGGAANGTGAGAVMGPNATAEEGAPS-----A 390  
 | | : : : | | | | : : | : : |  
 Db 466 RMNTPQGSFSTISQALCSQGI-----TTEYLTAMLPSSQRPKNHTKDFLTQYKLTKEQIA 520  
  
 Qy 391 AALATPDTLQGQCSAFVQ-----LWAGLQPILCGNNRTIEPEALRRGNMSSLGFT 440  
 : | : : : | | | : | | | |  
 Db 521 SKYGIPINTTFFCFSLYKDIINMPAGPVIWAFLKPMLLG----- 559

Qy 441 SKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVWLNIS-AE 499  
: ||||| : : | : : : | : : |

Db 560 -----RILYAPYNPVTKAIMEKSNVTLRQLAELREKSQEWMDKSPLF 601

Qy 500 IRSF--LEQG--RLQQHLR--WLQQYVA-ELRLHPEALNLSLDELPPALRQDNFSLPSGM 552  
: | | | | : | : | : | : | : | | : : | : :

Db 602 MNSFHLLNQAI PMLQNTLRNPFVQVFKFSVGLDAVELLKQIDEL-DILR---LKLENNI 657

Qy 553 ALLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQA-----Y 598  
: : ||:| : : ||:| : : ||:| : : ||:| : : ||:| : : ||:|

Db 658 DIIDQINT-----LSSLTVNI-----SSCVLYDRIQAAKTIDEMEREAKRLY 699

Qy 599 QDNVTVFASVIFQTRKDGS-----LPPHVHYKIRQNSSFTEKTNEIRRAYWRPG 647  
: | : | |||: : | ||| : | || : : | : | | ||

Db 700 KSN-ELFGSVIFKLPSNRSWHRGYDSGNVFLPPVIKYTIRMSLKTAQTTRSLRTKIWAPG 758

Qy 648 PNTGGREFYFLYG--FVWIQDMMERAIIDTFVGHDVVEPGSYVQMFYPCYTRDDFLFVIE 705  
| : : || | : : || : ||||: | : | || ||||: : ||:| :

Db 759 PHNSPSHNQIYGRAFIYLDQSIERAIIELOTGRNSQEIYVQVQAI PYPFCMKDNFLT SVS 818

Qy 706 HMMPLCMVISWVYSVAMTIQHIVA EKEHRLKEVMKTMGLNNAVHWVAWFITGFVQLSISV 765  
: : ||: : : ||: | : : | ||: || | || ||||: | : || | : : :

Db 819 YSLPIVLMVAWVVFIAAFVKKLVEYKDLRLHEYMKMMGVNSCSHFFAWLIESVGFLLVTI 878

Qy 766 TALTAILKYGQVLMHSHVVIWFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIFYLS 825  
| ||||: | : : | : ||: | : : | : ||: : : | : | : ||: :

Db 879 VILIIILKFGNILPKTNGFILFLYFSDYSFSVIAMSYLISVFFNNTNIAALIGSLIYIIA 938

Qy 826 YVPYMYVAIREEVAHDKITAFEKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSP 885  
: ||: : | : : : | ||: || | ||: | || | : ||: || ||

Db 939 FFPFIVLVTVE----NELSYVLKVFMSSLSPTAFSYASQYIARYEEQGIGLQWENMYTSP 994

Qy 886 VEGDDFNLLAVTMLMVDVAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYW---LGSGR 942  
| : | : : : | : : | : ||: | || ||: ||||: || |

Db 995 VQDDTTSFGWLCCILILADSEFIYFLIAWYVRNVFPPTYGMAAPWYFPILPYSYWKERFGCAE 1054

Qy 943 TEAWEWSWPWARTPRLS-----VMEEDQACAMESRRFEETRGMEEPTHLPVVCVD 994  
: | | : : | : | : : | | | : | :

Db 1055 VK-----PEKSNGLMFTNIMMQNTNPSASPEYMF--SSNIEPEPKDLTVGVALH 1101

Qy 995 KLTKVYKDDKKLALNKLSLNLYENQVVSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGH 1054  
: ||: | | : : : | : || || : | || ||||| : : |||| : : ||

Db 1102 GVTKIY--GSKVAVDNLNLNFYEGHITSLLGPNAGKTTTISMLTGLFGASAGTIFVYGK 1159

Qy 1055 DIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSRLK--SMAQEEIRREMDKMIEDLE 1112  
||: ||: : ||||: | ||: || || : || | : | : : : | : : : |

Db 1160 DIKTDLHTVRKNMGVCMQHDVLF SYLTKEHLLLYGSIKVPHWTKKQLHEEVKRTLKDTG 1219

Qy 1113 LSNKRHSILVQTLSGGMKRKLSVAIAFVGGSRAILDEPTAGVDPYARRAIWDLILKYKPG 1172  
| : || | ||||| ||||: || : |||| : ||||: |||| : ||: ||||: | |

Db 1220 LYSHRHKRVGTLSGGMKRKLSISIALIGGSRVVILDEPSTGVDPCSRRSIWDVISKNKTA 1279

Qy 1173 RTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPLFLKGTYG DGYRLTLVKRPAEPGGPQ 1232  
|||: ||||: ||||: | |||| : | : |||| : || : |||| || | : :

Db 1280 RTIILSTHHLDEAEVLSDRIAFLEQGGRLCCGS PFYLKEAFGDGYHLTLTKK-----K 1332



Qy 1233 EPGLASSPPGRAPLSSCSELQVSQFIRKHAVASCLLVSDTSTELSYILPSEAAK-KGA FER 1291  
 | | : : : | : | : | : | | | | : | | :  
 Db 1333 SPNLNAN-----AVCDTMAVTAMIQSHLPEAYLKEDIGGELVYVLPFPSTKVSGAYLS 1385

Qy 1292 LFQHLERSLDALHLSSFGMLDITTEEVFLKVSEEDQSLENSEADVKE SRKDVLPGAEGPA 1351  
 | : | : : | : : | : | : | : | : | : | : | : | :  
 Db 1386 LLRALDNGMGDLNIGCYGISDTTVEEVFLNLTKESQ--KNSAMSLE----- 1429

Qy 1352 SGEGHAGNLARCSELTQSQASLQSASSVGSARGDEGAGYTDVYGDYRPLFDNPQDPDNVS 1411  
 | | : : | : : : | : : : | : : : | : : : | : : :  
 Db 1430 -----HLTQKKIGNSNANGIST-----PDDL S 1451

Qy 1412 LQ-----EVEAEALSRVGQSGSRKLDGGWLKVQRQFHGLLVKRFHCARRNSKALFSQILLPA 1466  
 : : : : | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 1452 VSSSNFTDRDDKILTR----GERLDGFGLLKKIMAILIKRFHHTRRNWKGLIAQVILPI 1507

Qy 1467 FFVCVAMTVALSVP EIGDLPPLVLSPSQYHNYTQPRGNFIPYANEERREYRLRLSPDASP 1526  
 | | | : | : : | : | : | : | : | : | : | : | : | : | :  
 Db 1508 VEVTTAMGLGTLRNSNSYPEIQISPSLYG--TSEQTAF--YANYH-----PST 1552

Qy 1527 QQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPL 1586  
 : | | | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 1553 EALVSAMWDFPGI-----DNMCLNT----- 1572

Qy 1587 SNFVPPPPSPAPSDSPASPEDDLQAWNVS LPPTAGPEMWT SAPSLPRLVREPVR----CT 1642  
 | | : : | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 1573 -----SDLQCLNKDSLEKWNTS-----GEPITNFGVCS 1600

Qy 1643 CSAQGTGFSCPSSVGGHPPQMRVVTGDILTDTGHNVSEYLLFTSDRFLHRYGAITFGN 1702  
 | | | | | | | | | : : : : | : | : | : | : | : | : | : | :  
 Db 1601 CSENVQ--ECP-KFNYSPPHRRYSSQVIYNLTGQRVENYLISTANEFVQKRYGGWSFGL 1657

Qy 1703 VL-----KSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRA 1754  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 1658 PLTKDLRFDTITGV PAN-----RTLAKVWYDPEGYHSLPAYLNSLNNFLLRV 1703

Qy 1755 NLPKSKGNPAAYGITVTNHP----MNKTSASLSLDYLLQGTDVVI AFIIVAMSFVPASF 1810  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 1704 NM--SKYDAARHGIIMYSHPYPGVQDQE QATIS-----SLIDILVALSILMGYSVTTASF 1756

Qy 1811 VVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPATCCVILFVFDLPAYTSP 1870  
 | : : | | | | | | : | : | : | : | : | : | : | : | : | :  
 Db 1757 VTYVVREHQTKAKQLQHISGIGVTCYVVTNFIYDMVFYLVPAFSIGIIAIFKLPAFYSE 1816

Qy 1871 TNFPAVLSLFLLYGWSITPIMYPASFWEVPSSAYVFLIVINLFIGITA----TVATFLL 1926  
 | | | | | : : | : | : : | : | : | : | : | : | : | :  
 Db 1817 NNLGAVSLLLLLFGYATFSWMYLLAGLFHETGMAFITYVCVNLFFGINSIVLSVVFYLS 1876

Qy 1927 QLF EHDKDLKVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFEW 1986  
 : : | : : : | | | | : | : | : : : : : : | : | : | :  
 Db 1877 KEKPNDPTLELISETLKRIFLI FPQFCFGYGLIELSQQQSVLDFLKAYG-VEYPNETFEM 1935

Qy 1987 DIVTRGLVAMAVEGVVGFLLTIMCQYNFLR-----RPQ RMPVSTKPVEDD VDVASERQR 2040  
 : : | : : | : | : : : : | : : : : | : | : | : | :  
 Db 1936 NKLGAMEFVALVSQGT MFFSLRLLINESLIKKLRLFFRKFNSSHVRETIDEDEDVRAERLR 1995

Qy 2041 VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLG VNGAGKTSTFKM 2100

Db	1996	VESGAAEFDLVQLYCLTKTYQLIH-KKIIAVNNISIGIPAGECFGLLGVNGAGKTTIFKM	2054
Qy	2101	LTGD-----ESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQ	2150
Db	2055	LTGDIIPSSGNILIRNKTGSLGHVDSHSSL-----VGYPQEDALDDLVTVEEHLY	2105
Qy	2151	LYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFL	2210
Db	2106	FYARVHGIPEKDIKETVHKLLRRLHLMFPKDRATSMCSYGTKRKLSTALALIGKPSILL	2165
Qy	2211	DEPTTGMDPKARRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSI	2270
Db	2166	DEPSSGMDPKSKRHLWKIISEEVQNKCSVILTSHSMEECEALCTRLAIMVNGKFQCIGSL	2225
Qy	2271	QHLKNRFGDGYMITVRTKSSQ-SVKDVVRFFNRNFPPEAMLKERHHTKVQYQLKSEHISLA	2329
Db	2226	QHIKSRFGRGFTVKVHLKNNKVTMETLTKEFMQLHFPKTYLKDQHLSMLEYHVPVTTAGGVA	2285
Qy	2330	QVFSKMEQVSGVLGIEDYSVSQTTLDNVFNFAKKQSDNLEQQETEPSPALQSPL	2384
Db	2286	NIFDLLETNKTALNITNFLVSQTTLEEVFINFAKDQ----KSYETADTSSQGSTI	2336

Search completed: September 1, 2004, 10:57:10  
Job time : 242 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2004, 10:34:46 ; Search time 39 Seconds  
(without alignments)  
3252.382 Million cell updates/sec

Title: US-10-088-467-2  
Perfect score: 12668  
Sequence: 1 MGFLHQLQLLLWKNVTLKRR.....GLISFEEERAQLSFNTDTLC 2436

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	12658	99.9	2436	1 ABC2_HUMAN	Q9bzc7 homo sapien
2	11349	89.6	2434	1 ABC2_MOUSE	P41234 mus musculu
3	4233.5	33.4	2261	1 ABC1_HUMAN	O95477 homo sapien
4	4195.5	33.1	2261	1 ABC1_MOUSE	P41233 mus musculu
5	3875.5	30.6	2273	1 ABCR_HUMAN	P78363 homo sapien
6	2622	20.7	1704	1 ABC3_HUMAN	Q99758 homo sapien
7	1515	12.0	1704	1 CED7_CAEEL	P34358 caenorhabdi
8	405	3.2	330	1 DRRA_STRPE	P32010 streptomyce
9	382	3.0	340	1 NOD1_RHILO	P23703 rhizobium l
10	380	3.0	304	1 NOD1_RHIS3	P72335 rhizobium s
11	380	3.0	343	1 NOD1_RHISN	P55476 rhizobium s
12	379	3.0	347	1 NOD1_RHIGA	P50332 rhizobium g
13	367.5	2.9	335	1 NDI2_RHIME	Q8gnh6 rhizobium m
14	365.5	2.9	355	1 NDI1_RHIME	O52618 rhizobium m
15	354	2.8	578	1 YBHF_ECOLI	P75776 escherichia
16	348	2.7	306	1 NOD1_BRAJA	P26050 bradyrhizob
17	346	2.7	311	1 NOD1_RHILV	P08720 rhizobium l

18	344.5	2.7	894	1	YHIH_ECOLI	P37624	escherichia
19	332.5	2.6	308	1	NOSF_PSEST	P19844	pseudomonas
20	327.5	2.6	305	1	NODI_BRASS	Q9z3i3	bradyrhizob
21	324.5	2.6	308	1	YADG_ECOLI	P36879	escherichia
22	323	2.5	306	1	BCRA_BACLI	P42332	bacillus li
23	321.5	2.5	262	1	YA23_METJA	Q58429	methanococc
24	317.5	2.5	1280	1	MDR1_HUMAN	P08183	homo sapien
25	316.5	2.5	1321	1	MDR1_CAEEL	P34712	caenorhabdi
26	315.5	2.5	267	1	YATR_BACPF	P26946	bacillus ps
27	312.5	2.5	1499	1	CDR2_CANAL	P78595	candida alb
28	307.5	2.4	354	1	Y415_SYNY3	P22040	synechocyst
29	303	2.4	241	1	YHBG_HAEIN	P45073	haemophilus
30	301.5	2.4	335	1	Y7I9_ANASP	Q05067	anabaena sp
31	301.5	2.4	1276	1	MDR2_CRIGR	P21449	cricetulus
32	296	2.3	1501	1	CDR1_CANAL	P43071	candida alb
33	295.5	2.3	305	1	YHCH_BACSU	P54592	bacillus su
34	295.5	2.3	320	1	NODI_AZOCA	Q07756	azorhizobi
35	294	2.3	1276	1	MDR1_CRIGR	P21448	cricetulus
36	293	2.3	343	1	METN_SALTY	Q8zrm9	salmonella
37	292	2.3	1276	1	MDR1_MOUSE	P06795	mus musculu
38	289.5	2.3	380	1	OPCA_BACSU	O34992	bacillus su
39	287	2.3	343	1	METN_ECOL6	Q8x7z9	escherichia
40	286	2.3	343	1	METN_ECOLI	P30750	escherichia
41	286	2.3	343	1	METN_SALTI	Q8z990	salmonella
42	284	2.2	1276	1	MDR3_MOUSE	P21447	mus musculu
43	283.5	2.2	1490	1	CDR4_CANAL	O74676	candida alb
44	283	2.2	300	1	YCBN_BACSU	P42246	bacillus su
45	282	2.2	339	1	Y4FO_RHISN	P55453	rhizobium s

#### ALIGNMENTS

##### RESULT 1

##### ABC2\_HUMAN

ID ABC2\_HUMAN STANDARD; PRT; 2436 AA.  
AC Q9BZC7;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE ATP-binding cassette, sub-family A, member 2 (ATP-binding cassette  
DE transporter 2) (ATP-binding cassette 2).  
GN ABCA2 OR ABC2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=11178988;  
RA Kaminski W.E., Piehler A., Pullmann K., Porsch-Oezcueruemez M.,  
RA Duong C., Bared G.M., Buchler C., Schmitz G.;  
RT "Complete coding sequence, promoter region, and genomic structure of  
RT the human ABCA2 gene and evidence for sterol-dependent regulation in  
RT macrophages."  
RL Biochem. Biophys. Res. Commun. 281:249-258(2001).  
CC -!- FUNCTION: Probable transporter, its natural substrate has not been

CC found yet. May have a role in macrophage lipid metabolism and  
CC neural development.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCA subfamily.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AF327705; AAK14335.1; -.  
DR EMBL; AF327658; AAK14335.1; JOINED.  
DR EMBL; AF327659; AAK14335.1; JOINED.  
DR EMBL; AF327660; AAK14335.1; JOINED.  
DR EMBL; AF327661; AAK14335.1; JOINED.  
DR EMBL; AF327662; AAK14335.1; JOINED.  
DR EMBL; AF327663; AAK14335.1; JOINED.  
DR EMBL; AF327664; AAK14335.1; JOINED.  
DR EMBL; AF327665; AAK14335.1; JOINED.  
DR EMBL; AF327666; AAK14335.1; JOINED.  
DR EMBL; AF327667; AAK14335.1; JOINED.  
DR EMBL; AF327668; AAK14335.1; JOINED.  
DR EMBL; AF327669; AAK14335.1; JOINED.  
DR EMBL; AF327670; AAK14335.1; JOINED.  
DR EMBL; AF327671; AAK14335.1; JOINED.  
DR EMBL; AF327672; AAK14335.1; JOINED.  
DR EMBL; AF327673; AAK14335.1; JOINED.  
DR EMBL; AF327674; AAK14335.1; JOINED.  
DR EMBL; AF327675; AAK14335.1; JOINED.  
DR EMBL; AF327676; AAK14335.1; JOINED.  
DR EMBL; AF327677; AAK14335.1; JOINED.  
DR EMBL; AF327678; AAK14335.1; JOINED.  
DR EMBL; AF327679; AAK14335.1; JOINED.  
DR EMBL; AF327680; AAK14335.1; JOINED.  
DR EMBL; AF327681; AAK14335.1; JOINED.  
DR EMBL; AF327682; AAK14335.1; JOINED.  
DR EMBL; AF327683; AAK14335.1; JOINED.  
DR EMBL; AF327684; AAK14335.1; JOINED.  
DR EMBL; AF327685; AAK14335.1; JOINED.  
DR EMBL; AF327686; AAK14335.1; JOINED.  
DR EMBL; AF327687; AAK14335.1; JOINED.  
DR EMBL; AF327688; AAK14335.1; JOINED.  
DR EMBL; AF327689; AAK14335.1; JOINED.  
DR EMBL; AF327690; AAK14335.1; JOINED.  
DR EMBL; AF327691; AAK14335.1; JOINED.  
DR EMBL; AF327692; AAK14335.1; JOINED.  
DR EMBL; AF327693; AAK14335.1; JOINED.  
DR EMBL; AF327694; AAK14335.1; JOINED.  
DR EMBL; AF327695; AAK14335.1; JOINED.  
DR EMBL; AF327696; AAK14335.1; JOINED.  
DR EMBL; AF327697; AAK14335.1; JOINED.  
DR EMBL; AF327698; AAK14335.1; JOINED.  
DR EMBL; AF327699; AAK14335.1; JOINED.  
DR EMBL; AF327700; AAK14335.1; JOINED.

DR EMBL; AF327701; AAK14335.1; JOINED.  
 DR EMBL; AF327702; AAK14335.1; JOINED.  
 DR EMBL; AF327703; AAK14335.1; JOINED.  
 DR EMBL; AF327704; AAK14335.1; JOINED.  
 DR Genew; HGNC:32; ABCA2.  
 DR MIM; 600047; -.  
 DR GO; GO:0016021; C:integral to membrane; NAS.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; NAS.  
 DR GO; GO:0006629; P:lipid metabolism; NAS.  
 DR GO; GO:0006810; P:transport; NAS.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 2.  
 DR ProDom; PD000006; ABC\_transporter; 2.  
 DR SMART; SM00382; AAA; 2.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
 KW ATP-binding; Transport; Transmembrane; Repeat; Glycoprotein.  
 FT TRANSMEM 21 40 POTENTIAL.  
 FT TRANSMEM 706 728 POTENTIAL.  
 FT TRANSMEM 749 771 POTENTIAL.  
 FT TRANSMEM 786 808 POTENTIAL.  
 FT TRANSMEM 813 835 POTENTIAL.  
 FT TRANSMEM 850 872 POTENTIAL.  
 FT TRANSMEM 892 914 POTENTIAL.  
 FT TRANSMEM 1793 1815 POTENTIAL.  
 FT TRANSMEM 1846 1865 POTENTIAL.  
 FT TRANSMEM 1875 1897 POTENTIAL.  
 FT TRANSMEM 1904 1926 POTENTIAL.  
 FT TRANSMEM 1988 2010 POTENTIAL.  
 FT NP\_BIND 1025 1032 ATP (POTENTIAL).  
 FT NP\_BIND 2088 2095 ATP (POTENTIAL).  
 FT CARBOHYD 14 14 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 169 169 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 369 369 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 380 380 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 545 545 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 601 601 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1409 1409 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1497 1497 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1550 1550 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1558 1558 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1678 1678 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1776 1776 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2055 2055 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 2436 AA; 269971 MW; 9E6688D8615DE06D CRC64;

Query Match 99.9%; Score 12658; DB 1; Length 2436;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2435; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 MGFLHQQLLLWKNVTLKRRSPWVLAFEIFIPLVLFILLGLRQKKPTISVKEVPFYTAA 60
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MGFLHQQLLLWKNVTLKRRSPWVLAFEIFIPLVLFILLGLRQKKPTISVKEVPFYTAA 60

Qy     61 PLTSAGILPVMQSLCPDQORDEFGFLQYANSTVTQLLERLDRVVEEGNLFDPARPSLGSE 120
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PLTSAGILPVMQSLCPDQORDEFGFLQYANSTVTQLLERLDRVVEEGNLFDPARPSLGSE 120

Qy    121 LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSLPNSTAQAL 180
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSLPNSTAQAL 180

Qy    181 LAARVDPPEVYHLLFGPSSALDSQSGLHKQEPWSRLGGNPLFRMEELLAPALLEQLTC 240
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 LAARVDPPEVYHLLFGPSSALDSQSGLHKQEPWSRLGGNPLFRMEELLAPALLEQLTC 240

Qy    241 TPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARFSGLSAELRNQLDVAKVSQQ 300
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARFSGLSAELRNQLDVAKVSQQ 300

Qy    301 GLDAPNGSDSSPQAPPPRRLQALLGDLLDAQKVLQDQVDVLSALALLPQGACTGRTPGPP 360
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GLDAPNGSDSSPQAPPPRRLQALLGDLLDAQKVLQDQVDVLSALALLPQGACTGRTPGPP 360

Qy    361 ASGAGGAANGTGAGAVMGPNATAEEGAPSAALATPDTLQGQCSAFVQLWAGLQPILCGN 420
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ASGAGGAANGTGAGAVMGPNATAEEGAPSAALATPDTLQGQCSAFVQLWAGLQPILCGN 420

Qy    421 NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF 480
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKILYAPAGSEVDRVILKANETF 480

Qy    481 AFVGNVTHYAQVWLNISAEIRSFLEQGRLLQHLRWLQQYVAELRLHPEALNLSLDELPPA 540
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AFVGNVTHYAQVWLNISAEIRSFLEQGRLLQHLRWLQQYVAELRLHPEALNLSLDELPPA 540

Qy    541 LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPEESIVNYTLNQAYQD 600
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 LRQDNFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPEESIVNYTLNQAYQD 600

Qy    601 NVTVFASVIFQTRKDGSLLPPHVKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF 660
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 NVTVFASVIFQTRKDGSLLPPHVKIRQNSSFTEKTNEIRRAYWRPGPNTGGRFYFLYGF 660

Qy    661 VWIQDMMERAIIDTFVGHDVVEPGSYVQMFYPYCYTRDDFLFVIEHMMPLCMVISWVYSV 720
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 VWIQDMMERAIIDTFVGHDVVEPGSYVQMFYPYCYTRDDFLFVIEHMMPLCMVISWVYSV 720

Qy    721 AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAWFITGFVQLSISVTALTAILKYQVLMH 780
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAWFITGFVQLSISVTALTAILKYQVLMH 780
```

Qy	781	SHVVIIWLF LAVYAVATIMFCFLSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEV AH	840
Db	781	SHVVIIWLF LAVYAVATIMFCFLSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEV AH	840
Qy	841	DKITAF EKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLL LAVTML	900
Db	841	DKITAF EKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLL LAVTML	900
Qy	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSV	960
Db	901	MVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWEWSWPWARTPRLSV	960
Qy	961	MEEDQACAMESRRFEETRGMEEEP THLPLVVCVDKLT KVYKDDKKLALNKLSLNLYENQV	1020
Db	961	MEEDQACAMESRRFEETRGMEEEP THLPLVVCVDKLT KVYKDDKKLALNKLSLNLYENQV	1020
Qy	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR L	1080
Db	1021	VSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDR L	1080
Qy	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAFVG	1140
Db	1081	TVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAFVG	1140
Qy	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Db	1141	GSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKL	1200
Qy	1201	KCCGSPLFLKGT YGDYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Db	1201	KCCGSPLFLKGT YGDYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRK	1260
Qy	1261	HVASCLLVSDTSTELSYILPSEA AKKGA FERLFQHLERSLDALHLSSFG LMDTTLEEVFL	1320
Db	1261	HVASCLLVSDTSTELSYILPSEA AKKGA FERLFQHLERSLDALHLSSFG LMDTTLEEVFL	1320
Qy	1321	KVSEEDQSLENSEADV KESRKDVLP GAEGPASGEGHAGNLARCSELTQSQASLQSASSVG	1380
Db	1321	KVSEEDQSLENSEADV KESRKDVLP GAEGPASGEGHAGNLARCSELTQSQASLQSASSVG	1380
Qy	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Db	1381	SARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRKLDGGWLKVRQF	1440
Qy	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYHNYTQ	1500
Db	1441	HGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYHNYTQ	1500
Qy	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Db	1501	PRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLS	1560
Qy	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPEDLQAWNVS LPTTA	1620
Db	1561	SGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPEDLQAWNVS LPTTA	1620



Qy	1621	GPBMWTSAPSLPRLVREPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILTDITGHNVS	1680
Db	1621	GPBMWTSAPSLPRLVREPVRCTCSAQGTGFSCPSVGGHPPQMRVVTGDILTDITGHNVS	1680
Qy	1681	EYLLFTSDRFLRHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Db	1681	EYLLFTSDRFLRHRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKGYHSM	1740
Qy	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFII	1800
Db	1741	PTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDVVIAIFII	1800
Qy	1801	VAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWMLNYLVPATCCVIIIF	1860
Db	1801	VAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWMLNYLVPATCCVIIIF	1860
Qy	1861	VFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWEFVPSSAYVFLIVINLFIGITAT	1920
Db	1861	VFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFWEFVPSSAYVFLIVINLFIGITAT	1920
Qy	1921	VATFLLQLFEHDKDLKVVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKM	1980
Db	1921	VATFLLQLFEHDKDLKVVNSYLKSCFLIFPNYNLGHGLMEMAYNEYINEYYAKIGQFDKM	1980
Qy	1981	KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDVDVASERQR	2040
Db	1981	KSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDVDVASERQR	2040
Qy	2041	VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRCLCLGVRPGECFGLLGVNGAGKTSTFKM	2100
Db	2041	VLRGDADNDMVKIENLTKVYKSRKIGRILAVDRCLCLGVRPGECFGLLGVNGAGKTSTFKM	2100
Qy	2101	LTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW	2160
Db	2101	LTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISW	2160
Qy	2161	KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK	2220
Db	2161	KDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPK	2220
Qy	2221	ARRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG	2280
Db	2221	ARRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDG	2280
Qy	2281	YMITVRTKSSQSVKDVVRFFNRNFPPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSG	2340
Db	2281	YMITVRTKSSQSVKDVVRFFNRNFPPEAMLKERHHTKVQYQLKSEHISLAQVFSKMEQVSG	2340
Qy	2341	VLGIEDYSVSQTTLDNVFNFAKKQSDNLEQQETEPSPALQSPLGCLLSLLRPRSAPTEL	2400
Db	2341	VLGIEDYSVSQTTLDNVFNFAKKQSDNLEQQETEPSPALQSPLGCLLSLLRPRSAPTEL	2400
Qy	2401	RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC	2436
Db	2401	RALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC	2436

RESULT 2

ABC2\_MOUSE

ID ABC2\_MOUSE STANDARD; PRT; 2434 AA.  
AC P41234;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE ATP-binding cassette, sub-family A, member 2 (ATP-binding cassette  
DE transporter 2) (ATP-binding cassette 2).  
GN ABCA2 OR ABC2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A., AND REVISIONS.  
RC STRAIN=DBA/2;  
RA Chimini G.;  
RL Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE OF 964-2434 FROM N.A.  
RC STRAIN=DBA/2; TISSUE=Macrophage;  
RX MEDLINE=94375008; PubMed=8088782;  
RA Luciani M.F., Denizot F., Savary S., Mattei M.-G., Chimini G.;  
RT "Cloning of two novel ABC transporters mapping on human chromosome  
RT 9.";  
RL Genomics 21:150-159(1994).  
CC -!- FUNCTION: Probable transporter, its natural substrate has not been  
CC found yet. May have a role in macrophage lipid metabolism and  
CC neural development.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- TISSUE SPECIFICITY: Widely expressed in adult tissues. Highest  
CC levels are found in brain and pregnant uterus.  
CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCA subfamily.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X75927; CAA53531.2; -.  
DR MGD; MGI:99606; Abca2.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR Pfam; PF00005; ABC\_tran; 2.  
DR ProDom; PD000006; ABC\_transporter; 2.  
DR SMART; SM00382; AAA; 2.  
DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
KW ATP-binding; Transport; Transmembrane; Repeat; Glycoprotein.  
FT TRANSMEM 21 40 POTENTIAL.  
FT TRANSMEM 705 727 POTENTIAL.  
FT TRANSMEM 748 770 POTENTIAL.  
FT TRANSMEM 780 802 POTENTIAL.

FT	TRANSMEM	809	831	POTENTIAL.
FT	TRANSMEM	1793	1815	POTENTIAL.
FT	TRANSMEM	1846	1865	POTENTIAL.
FT	TRANSMEM	1875	1897	POTENTIAL.
FT	TRANSMEM	1904	1926	POTENTIAL.
FT	NP_BIND	1024	1031	ATP (POTENTIAL).
FT	NP_BIND	2088	2095	ATP (POTENTIAL).
FT	CARBOHYD	14	14	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	89	89	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	168	168	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	173	173	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	305	305	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	368	368	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	379	379	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	420	420	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	432	432	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	476	476	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	484	484	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	494	494	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	530	530	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	548	548	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	589	589	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	599	599	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	627	627	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1408	1408	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1496	1496	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1549	1549	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1557	1557	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1613	1613	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1678	1678	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1776	1776	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2055	2055	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	2434	AA; 270582 MW; 3CEDD48ED5692005	CRC64;

Query Match 89.6%; Score 11349; DB 1; Length 2434;  
 Best Local Similarity 90.6%; Pred. No. 0;  
 Matches 2217; Conservative 51; Mismatches 155; Indels 24; Gaps 9;

Qy	1	MGFLHQLQLLLWKNVTLKRRSPWVLA FEIFIPLVLF FILLGLRQKKPTISVKEVPFYTAA	60
Db	1	MGFLHQLQLLLWKNVTLKRRSPWVLA FEIFIPLVLF FILLGLRQKKPTISVKEA-FYTAA	59
Qy	61	PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLDRVVEEGNLFDPARPSLGSE	120
Db	60	PLTSAGILPVMQSLCPDGQRDEFGFLQYANSTVTQLLERLHRVVEEGNLFDPVRPSLGSE	119
Qy	121	LEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNL SLPNSTAQAL	180
Db	120	LEALRQRLEALSSGPGTWESH SARPAVSSFSLDSVARDQRELWRFLMQNL SLPNSTAQAL	179
Qy	181	LAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEELL LAPALLEQLTC	240
Db	180	LAARVDPSEVYRLLFGPLPDLGKLGFLRKQEPWSRLGSNPLLQMEELL LAPALLEQLTC	239
Qy	241	TPGSGELGRILTV PESQK GALQGYRDAVCSGQAAARARFSGLSAELRNQLD VAKVSQQL	300
Db	240	APGSGELGRILTMP EGHQVDLQGYRDAVCSGQATARAQRFSDLAAELRNQLD TAKIAQQL	299

Qy	301	GLDAPNGSDSSPQAPPPRRLLQALLGDLLDAQKVLQDQDVL	360
Db	300	GFDVPNGSDPQFPQAPSPQSLPALLGDLLDAQKLLQDQDVL	359
Qy	361	ASGAGGAANGTGAGAVMGPNATAEEGAPSAALATPDTLQGC	420
Db	360	ASSLNLGLANSTGIGANSNSTTVEEGTQSPVSPASPTLQGC	419
Qy	421	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKIL	480
Db	420	NRTIEPEALRRGNMSSLGFTSKEQRNLGLLVHLMTSNPKIL	479
Qy	481	AFVGNVTHYAQVWLNISAEIRSFLEQGRLLQHLRWLQQYV	540
Db	480	AFVGNVTHYAQVWLNISTEIRSFLEQGRLLQHLRWLQQYV	539
Qy	541	LRQDNFSLPSGMALLQQLDITDNAACGWIQFMSKVSVDIF	600
Db	540	LRQD-FSLPNGTALLQQLDITDNAACGWIQFMSKVSVDIF	598
Qy	601	NVTVFASVIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIR	660
Db	599	NVTVFASVIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIR	658
Qy	661	VWIQDMMERAIIDTFVGHDVVEPGSYVQMFYPYCYTRDD	720
Db	659	RLDQDMMERAIINTFVGHDVVEPGNYVQMFYPYCYTRDD	718
Qy	721	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAWFITGFVQ	780
Db	719	AMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAWFITGFVQ	778
Qy	781	SHVVI IWLFLAVYAVATIMFCFLVSVLYSKAKLASACGG	840
Db	779	SHVLI IWLFLAVYAVATIMFCFLVSVLYSKAKLASACGG	838
Qy	841	DKITAFEKCIASLMSTTAFG-LGSKYFALYEAVGVGIQWH	899
Db	839	DKITAFEKCIASRCPQQPLAWVPSTLHCMKWQEWASIQW	898
Qy	900	LMVDAVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSY-WL	950
Db	899	LMVDTVVYGVLTYWYIEAVHPGMYGLPRPWYSRYSPIGW	958
Qy	951	PWARTPRLSVMEEDQACAMESRRFEETRGMEEEPTHLPLV	1010
Db	959	LWRRI-----QACAMESRHFEETRGMEEEPTHLPLV	1009
Qy	1011	LSLNLYENQVVSFLGHNGAGKTTTMSILTGLFPPTSGSAT	1070
Db	1010	LSLNLYENQVVSFLGHNGAGKTTTMSILTGLFPPTSGSAT	1069
Qy	1071	PQHNVLFDRLTVEEHLWFYSRLKSMAQEEIRREMDKMIED	1130
Db	1070	PQHNVLFDRLTVEEHLWFYSRLKSMAQEEIRKETDKMIED	1129

Qy 1131 KLSVAIAFVGGSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGD 1190  
 |||||  
 Db 1130 KLSVAIAFVGGSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGD 1189

Qy 1191 RIAIISHGKLKCCGSPLFLKGTYG DGYRLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCS 1250  
 ||||| : |||||  
 Db 1190 RIAIISHGKLKCCGSPLFLKGAYXDGYRLTLVKQPAEPGTSQEPGLASSPSGCPRLSSCS 1249

Qy 1251 ELQVSQFIRKHAVASCLLVSDTSTELSYILPSEAAKKGA FERL FQH LERSLDALHLSSFGL 1310  
 | |||||  
 Db 1250 EPQVSQFIRKHAVASSLLVSDTSTELSYILPSEAVKKGA FERL FQQ LEHSLDALHLSSFGL 1309

Qy 1311 MDTTLEEVFLKVSEEDQSLENSEADV KESRKDVLP GAEGPASGEGHAGNLARCSELTQSQ 1370  
 ||||| : | |||||  
 Db 1310 MDTTLEEVFLKVSEEDQSLENSEADV KESRKDVLP GAEGLTAVGGQAGNLARCSELAQSQ 1369

Qy 1371 ASLQSASSVGSARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVGQGSRK 1430  
 ||||| : || || : | ||||| ||||| | ||| : |||||  
 Db 1370 ASLQSASSVGSARGEETGYSDGYGDYRPLFDNLQDPDNVSLQEAEMEALAQVGQGSRK 1429

Qy 1431 DGGWLKVRQFHGLLVKRFHCARRNSKALFSQILLPAFFVVCVAMTVALSVPEIGDLPPLVL 1490  
 : | ||| : |||||  
 Db 1430 EGWWLKMVRQFHGLLVKRFHCARRNSKALCSQILLPAFFVVCVAMTVALSVPEIGDLPPLVL 1489

Qy 1491 SPSQYHNYTQPRGNFI PYANEERREYRLRLSPDAS PQQLVSTFRLPSGVGATCVLKSPAN 1550  
 ||||| : |||||  
 Db 1490 SPSQYHNYTQPRGNFI PYANEERQEYRLRLSPDAS PQQLVSTFRLPSGVGATCVLKSPAN 1549

Qy 1551 GSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPED-L 1609  
 |||| |||||  
 Db 1550 GSLGPTLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPVXPEDSL 1609

Qy 1610 QAWNVS LPP TAGPEMWT SAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGD 1669  
 |||| : |||||  
 Db 1610 QAWNVS LPP TAGPETWT SAPSLPRLVHEPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGD 1669

Qy 1670 ILTDITGHN VSEYLLFTSDRFR LHRYGAITFGNVLSKIPASFGTRAPPMVRKIAVRRAAQ 1729  
 |||||  
 Db 1670 ILTDITGHN VSEYLLFTSDRFR LHRYGAITFGNVQKIPASFGARVPPMVRKIAVRRVAQ 1729

Qy 1730 VFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHMPMNKTSASLSLDYLLQ 1789  
 | |||||  
 Db 1730 VLYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHMPMNKTSASLSLDYLLQ 1789

Qy 1790 GTDVVIAIFIIVAMSFVPASFVVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYL 1849  
 ||||| : |||||  
 Db 1790 GTDVVIAIFIIVAMSFVPASFVVFLVAEKSTKAKHLQFVSGCNPVIYWLANYVWDMNLNYL 1849

Qy 1850 VPATCCVIIIFVFDLPAYTSPTNFP AVL SLFLLYGWSITPIMYPASFWEVPSSAYVFLI 1909  
 |||||  
 Db 1850 VPATCCVIIIFVFDLPAYTSPTNFP AVL SLFLLYGWSITPIMYPASFWEVPSSAYVFLI 1909

Qy 1910 VINLFIGITATVATFLLQLFEHDKDLKVNSYKSCFLIFPNYNLGHGLMEMAYNEYINE 1969  
 |||||  
 Db 1910 VINLFIGITATVATFLLQLFEHDKDLKVNSYKSCFLIFPNYNLGHGLMEMAYNEYINE 1969

Qy 1970 YYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRM PVSTKPVE 2029

```

Db      1970 YYAKIGQFDKMKSPFEWDIVTRGLVAMTVEGFVGFFLTIMCQYNFLRQPQRLPVSTKPVE 2029
Qy      2030 DDVDVASERQVRVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLG 2089
Db      2030 DDVDVASERQVRVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLG 2089
Qy      2090 NGAGKTSTFKMLTGDESTTGGEAFVNGHSLVKELLQVQQSLGYCPQCDALFDELTAREHL 2149
Db      2090 NGAGKTSTFKMLTGDESTTGGEAFVNGHSLVKDLLQVQQSLGYCPQFDALFDELTAREHL 2149
Qy      2150 QLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIF 2209
Db      2150 QLYTRLRGIPWKDEAQVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIF 2209
Qy      2210 LDEPTTGMDPKARRFLWNLIILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGS 2269
Db      2210 LDEPTTGMDPKARRFLWNLIILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGS 2269
Qy      2270 IQHLKNRFGDGYMITVRTKSSQSVKDVVRFFNRNFP EAMLERHHTKVQYQLKSEHISLA 2329
Db      2270 IQHLKNRFGDGYMITVRTKSSQNVKDVVRFFNRNFP EAMLERHHTKVQYQLKSEHISLA 2329
Qy      2330 QVFSKMEQVSGVLGIEDYSVSQTTLDNVFNFAKKQSDNLEQQETEPPSALQSPLGCLLS 2389
Db      2330 QVFSKMEQVVGVLGIEDYSVSQTTLDNVFNFAKKQSDNVEQQEAE-PSSLPSPLG-LLS 2387
Qy      2390 LLRPRSAPTELRALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2436
Db      2388 LLRPRPAPTELRALVADEPEDLDTEDEGLISFEEERAQLSFNTDTLC 2434

```

### RESULT 3

#### ABC1\_HUMAN

```

ID   ABC1_HUMAN      STANDARD;          PRT;   2261 AA.
AC   O95477; Q96S56; Q96T85; Q9NQV4; Q9UN06; Q9UN07; Q9UN08; Q9UN09;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   ATP-binding cassette, sub-family A, member 1 (ATP-binding cassette
DE   transporter 1) (ATP-binding cassette 1) (ABC-1) (Cholesterol efflux
DE   regulatory protein).
GN   ABCA1 OR ABC1 OR CERP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20345099; PubMed=10884428;
RA   Santamarina-Fojo S., Peterson K.M., Knapper C.L., Qiu Y.,
RA   Freeman L.A., Cheng J.-F., Osorio J., Remaley A.T., Yang X.-P.,
RA   Haudenschild C.C., Prades C., Chimini G., Blackmon E.E.,
RA   Francois T.L., Duverger N., Rubin E.M., Rosier M., Deneffe P.,
RA   Fredrickson D.S., Brewer H.B. Jr.;
RT   "Complete genomic sequence of the human ABCA1 gene: analysis of the
RT   human and mouse ATP-binding cassette A promoter.";

```

RL Proc. Natl. Acad. Sci. U.S.A. 97:7987-7992(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RA Schwartz K., Lawn R.M., Wade D.P.;  
 RT "ABCA1 gene expression and apoA-I-mediated cholesterol efflux are  
 RT regulated by LXR.";  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21251004; PubMed=11352567;  
 RA Qiu Y., Cavelier L., Chiu S., Yang X., Rubin E., Cheng J.-F.;  
 RT "Human and mouse ABCA1 comparative sequencing and transgenesis  
 RT studies revealing novel regulatory sequences.";  
 RL Genomics 73:66-76(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Tanaka A.R., Abe-Dohmae S., Arakawa R., Sadanami K., Kidera A.,  
 RA Kioka N., Amachi T., Yokoyama S., Ueda K.;  
 RT "A new topological model of functional human ABCA1-signal peptide  
 RT cleavage and glycosylation of a large extracellular domain.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE OF 21-2261 FROM N.A.  
 RX MEDLINE=99194549; PubMed=10092505;  
 RA Langmann T., Klucken J., Reil M., Liebisch G., Luciani M.F.,  
 RA Chimini G., Kaminski W.E., Schmitz G.;  
 RT "Molecular cloning of the human ATP-binding cassette transporter 1  
 RT (hABCL1): evidence for sterol-dependent regulation in macrophages.";  
 RL Biochem. Biophys. Res. Commun. 257:29-33(1999).  
 RN [6]  
 RP SEQUENCE OF 21-2261 FROM N.A.  
 RX MEDLINE=99364413; PubMed=10431238;  
 RA Rust S., Rosier M., Funke H., Real J., Amoura Z., Piette J.-C.,  
 RA Deleuze J.-F., Brewer H.B., Duverger N., Deneffe P., Assmann G.;  
 RT "Tangier disease is caused by mutations in the gene encoding  
 RT ATP-binding cassette transporter 1.";  
 RL Nat. Genet. 22:352-355(1999).  
 RN [7]  
 RP PHOSPHORYLATION OF SER-1042 AND SER-2054.  
 RX MEDLINE=22289331; PubMed=12196520;  
 RA See R.H., Caday-Malcolm R.A., Singaraja R.R., Zhou S., Silverston A.,  
 RA Huber M.T., Moran J., James E.R., Janoo R., Savill J.M., Rigot V.,  
 RA Zhang L.H., Wang M., Chimini G., Wellington C.L., Tafuri S.R.,  
 RA Hayden M.R.;  
 RT "Protein kinase A site-specific phosphorylation regulates ATP-binding  
 RT cassette A1 (ABCA1)-mediated phospholipid efflux.";  
 RL J. Biol. Chem. 277:41835-41842(2002).  
 RN [8]  
 RP VARIANTS HDLD2 THR-1091 AND 1893-GLU-ASP-1894 DEL.  
 RX MEDLINE=20001430; PubMed=10533863;  
 RA Marcil M., Brooks-Wilson A., Clee S.M., Roomp K., Zhang L.-H., Yu L.,  
 RA Collins J.A., van Dam M., Molhuizen H.O.F., Loubser O.,  
 RA Ouelette B.F.F., Sensen C.W., Fichter K., Mott S., Denis M.,  
 RA Boucher B., Pimstone S., Genest J. Jr., Kastelein J.J.P., Hayden M.R.;  
 RT "Mutations in the ABC1 gene in familial HDL deficiency with defective  
 RT cholesterol efflux.";

RL Lancet 354:1341-1346(1999).  
 RN [9]  
 RP VARIANTS HDLD1 ARG-597 AND ARG-1477, AND VARIANT HDLD2 LEU-693 DEL.  
 RX MEDLINE=99364411; PubMed=10431236;  
 RA Brooks-Wilson A., Marcil M., Clee S.M., Zhang L.-H., Roomp K.,  
 RA van Dam M., Yu L., Brewer C., Collins J.A., Molhuizen H.O.F.,  
 RA Loubser O., Ouelette B.F.F., Fichter K., Ashbourne-Excoffon K.J.D.,  
 RA Sensen C.W., Scherer S., Mott S., Denis M., Martindale D.,  
 RA Frohlich J., Morgan K., Koop B., Pimstone S., Kastelein J.J.P.,  
 RA Hayden M.R.;  
 RT "Mutations in ABC1 in Tangier disease and familial high-density  
 RT lipoprotein deficiency.";  
 RL Nat. Genet. 22:336-345(1999).  
 RN [10]  
 RP VARIANTS HDLD1 SER-590; SER-935 AND VAL-937, AND VARIANTS ALA-399 AND  
 RP MET-883.  
 RX MEDLINE=99364412; PubMed=10431237;  
 RA Bodzioch M., Orso E., Klucken J., Langmann T., Boettcher A.,  
 RA Diederich W., Drobnik W., Barlage S., Buechler C.,  
 RA Porsch-Oezcueromez M., Kaminski W.E., Hahmann H.W., Oette K.,  
 RA Rothe G., Aslanidis C., Lackner K.J., Schmitz G.;  
 RT "The gene encoding ATP-binding cassette transporter 1 is mutated in  
 RT Tangier disease.";  
 RL Nat. Genet. 22:347-351(1999).  
 RN [11]  
 RP VARIANTS HDLD1 ILE-929; ARG-597 AND ARG-1477, AND VARIANTS HDLD2  
 RP LEU-693 DEL; THR-1091; 1893-GLU-ASP-1894 DEL AND LEU-2150.  
 RX MEDLINE=20540002; PubMed=11086027;  
 RA Clee S.M., Kastelein J.J.P., van Dam M., Marcil M., Roomp K.,  
 RA Zwarts K.Y., Collins J.A., Roelants R., Tamasawa N., Stulc T.,  
 RA Suda T., Ceska R., Boucher B., Rondeau C., DeSouich C.,  
 RA Brooks-Wilson A., Molhuizen H.O.F., Frohlich J., Genest J. Jr.,  
 RA Hayden M.R.;  
 RT "Age and residual cholesterol efflux affect HDL cholesterol levels and  
 RT coronary artery disease in ABCA1 heterozygotes.";  
 RL J. Clin. Invest. 106:1263-1270(2000).  
 RN [12]  
 RP VARIANTS HDLD1 ASN-1289 AND HIS-1800.  
 RX MEDLINE=20171564; PubMed=10706591;  
 RA Brousseau M.E., Schaefer E.J., Dupuis J., Eustace B.,  
 RA Van Eerdewegh P., Goldkamp A.L., Thurston L.M., FitzGerald M.G.,  
 RA Yasek-McKenna D., O'Neill G., Eberhart G.P., Weiffenbach B.,  
 RA Ordovas J.M., Freeman M.W., Brown R.H. Jr., Gu J.Z.;  
 RT "Novel mutations in the gene encoding ATP-binding cassette 1 in four  
 RT tangier disease kindreds.";  
 RL J. Lipid Res. 41:433-441(2000).  
 RN [13]  
 RP VARIANT HDLD1 ASP-1046, VARIANT HDLD2 CYS-230, AND VARIANTS LYS-219;  
 RP ILE-825; MET-883 AND LYS-1587.  
 RX MEDLINE=20396633; PubMed=10938021;  
 RA Wang J., Burnett J.R., Near S., Young K., Zinman B., Hanley A.J.G.,  
 RA Connelly P.W., Harris S.B., Hegele R.A.;  
 RT "Common and rare ABCA1 variants affecting plasma HDL cholesterol.";  
 RL Arterioscler. Thromb. Vasc. Biol. 20:1983-1989(2000).  
 RN [14]  
 RP VARIANT HDLD1 TRP-587, AND VARIANT LEU-2168.  
 RX MEDLINE=21157002; PubMed=11257260;



RA Bertolini S., Pisciotto L., Seri M., Cusano R., Cantafora A.,  
 RA Calabresi L., Franceschini G., Ravazzolo R., Calandra S.;  
 RT "A point mutation in ABC1 gene in a patient with severe premature  
 RT coronary heart disease and mild clinical phenotype of Tangier  
 RT disease.";  
 RL Atherosclerosis 154:599-605(2001).  
 RN [15]  
 RP VARIANTS LYS-219; MET-883 AND ASP-1172.  
 RX MEDLINE=21157003; PubMed=11257261;  
 RA Brousseau M.E., Bodzioch M., Schaefer E.J., Goldkamp A.L., Kielar D.,  
 RA Probst M., Ordovas J.M., Aslanidis C., Lackner K.J.,  
 RA Bloomfield Rubins H., Collins D., Robins S.J., Wilson P.W.F.,  
 RA Schmitz G.;  
 RT "Common variants in the gene encoding ATP-binding cassette transporter  
 RT 1 in men with low HDL cholesterol levels and coronary heart disease.";  
 RL Atherosclerosis 154:607-611(2001).  
 RN [16]  
 RP VARIANT HDLD1 LEU-1506.  
 RX MEDLINE=21369429; PubMed=11476961;  
 RA Lapicka-Bodzioch K., Bodzioch M., Kruell M., Kielar D., Probst M.,  
 RA Kiec B., Andrikovics H., Boettcher A., Hubacek J., Aslanidis C.,  
 RA Suttorp N., Schmitz G.;  
 RT "Homogeneous assay based on 52 primer sets to scan for mutations of  
 RT the ABCA1 gene and its application in genetic analysis of a new  
 RT patient with familial high-density lipoprotein deficiency syndrome.";  
 RL Biochim. Biophys. Acta 1537:42-48(2001).  
 RN [17]  
 RP VARIANTS HDLD1 ASN-1289 AND TRP-2081, AND VARIANT LYS-219.  
 RX MEDLINE=21369433; PubMed=11476965;  
 RA Huang W., Moriyama K., Koga T., Hua H., Ageta M., Kawabata S.,  
 RA Mawatari K., Imamura T., Eto T., Kawamura M., Teramoto T., Sasaki J.;  
 RT "Novel mutations in ABCA1 gene in Japanese patients with Tangier  
 RT disease and familial high density lipoprotein deficiency with  
 RT coronary heart disease.";  
 RL Biochim. Biophys. Acta 1537:71-78(2001).  
 RN [18]  
 RP VARIANTS LYS-219; ALA-399; MET-771; PRO-774; ASN-776; ILE-825;  
 RP MET-883; ASP-1172; LYS-1587 AND CYS-1731.  
 RX MEDLINE=21138379; PubMed=11238261;  
 RA Clee S.M., Zwinderman A.H., Engert J.C., Zwarts K.Y.,  
 RA Molhuizen H.O.F., Roomp K., Jukema J.W., van Wijland M., van Dam M.,  
 RA Hudson T.J., Brooks-Wilson A., Genest J. Jr., Kastelein J.J.P.,  
 RA Hayden M.R.;  
 RT "Common genetic variation in ABCA1 is associated with altered  
 RT lipoprotein levels and a modified risk for coronary artery disease.";  
 RL Circulation 103:1198-1205(2001).  
 RN [19]  
 RP VARIANT HDLD1 THR-255, AND VARIANT ATHEROSCLEROSIS ASP-1611.  
 RX MEDLINE=21645894; PubMed=11785958;  
 RA Nishida Y., Hirano K., Tsukamoto K., Nagano M., Ikegami C., Roomp K.,  
 RA Ishihara M., Sakane N., Zhang Z., Tsujii K., Matsuyama A., Ohama T.,  
 RA Matsuura F., Ishigami M., Sakai N., Hiraoka H., Hattori H.,  
 RA Wellington C., Yoshida Y., Misugi S., Hayden M.R., Egashira T.,  
 RA Yamashita S., Matsuzawa Y.;  
 RT "Expression and functional analyses of novel mutations of ATP-binding  
 RT cassette transporter-1 in Japanese patients with high-density  
 RT lipoprotein deficiency.";



Db	536	IVFTGITPGSIELPHHVKYKIRMDIDNVERTNKIKDGYWDPGPRADPFEDMRYVWGGFAY	595
Qy	663	IQDMMERAIIDTFVGHDVVEPGSYVQMFPYPCYTRDDFLFVIEHMMPLCMVISWVYSVAM	722
Db	596	: ::      : :    :             :       :::   :	
Qy	723	LQDVVEQAIIRVLGTGE-KKTGVYMQQMPYPYVDDIFLRVMSRSMPLFMTLAWIYSVAV	654
Qy	783	TIQHIVAEKEHRLKEVMKTMGLNNVAHWVAWFITGFVQLSISVTALTAILKYQQLMHSH	782
Db	655	:              :    :: :   :  : :   :          :  :	
Qy	783	IIKGIVYEKEARLKETMRIMGLDNSILWFSWFISSLIPLLVSAGLLVVILKLGNLLPYSD	714
Qy	783	VVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIFYLSYVPYMYVAIREEVAHDK	842
Db	715	::: :: :      :   :   : :   :        :: :	
Qy	843	PSVVFVFLSVFAVVTILQCFLISTLFSRANLAACGGIIFTLYLPYVLC-----VAWQD	769
Qy	843	ITAFE-KCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVEGDDFNLLLAVTMLM	901
Db	770	:        :   :   : :  :          : : :	
Qy	902	YVGFTLKFIFASLLSPAFAFGFGCEYFALFEEQGIGVQWDNLFESPVEEDGFNLTTSVSMML	829
Qy	902	VDAVVYGILTWYIEAVHPGMYGLRPWPYFPLQKSXYWLGSGRTEAWESWPWARTPRLSVM	961
Db	830	: :: :           :               :     :   :	
Qy	962	FDTFLYGMTWYIEAVFPGQYGI PRPWYF PCTKSYW FGE---ESDEKSHPGSNQKRIS--	884
Qy	962	EEDQACAMESRRFEETRGMEEEPHTLPLVCVDKLTKVYKDDKKLALNKLSLNLYENQVV	1021
Db	885	:               :      :   : :~ :        :	
Qy	1022	---EIC-----MEEEPHTLKLGVSIQNLVKVYRDGMKVAVDGLALNFYEGQIT	929
Qy	1022	SFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLT	1081
Db	930	:        :      :  :	
Qy	1082	SFLGHNGAGKTTTMSILTGLFPPTSGTAYILGKDIRSEMSTIRQNLGVCQPQHNVLFDMLT	989
Qy	1082	VEEHLWFYSRLKSMAQEERREMDKMIEDLEL-SNKRHSVLQTLSSGGMKRKLSVAIAFVG	1140
Db	990	: : :    ::~:  ::   :    :         :     :	
Qy	1141	VEEHWFYARLKLSEKHVKAEMEQAQMALDVGLPSSKLKSKTSQLSSGMQRKLSVALAFVG	1049
Qy	1141	GSRAILDEPTAGVDPYARRAIWDLILKYKPGRITILLSTHHMDEADLLGDRIAIISHGKL	1200
Db	1050	: ~ :           :       : :      : :	
Qy	1201	GSKVILDEPTAGVDPYSRRGIWELLKYRQGRITILSTHHMDEADVLDRIAIISHGKL	1109
Qy	1201	KCCGSPLFLKGTGYDGYRLTLVKRPAEPG-----GPQEPGLAS	1238
Db	1110	:   :	
Qy	1239	CCVGSSLFLKNQLGTGYLTLVKKDVESSLSCRNSSSTVSYLKKEDSVSQSSSDAGLGS	1169
Qy	1239	SPPGRAPLSSCSELQVSQFIRKHVASCLLVSDTSTELSYILPSEAARKGAFERLFQHLER	1298
Db	1170	:       :        : :       :      ::	
Qy	1299	DHESDTLTIDVS--AISNLIRKHVSEARLVEDIGHELTYVLPYEAAKEGAFVELFHEIDD	1227
Qy	1299	SLDALHLSSFGLMDTTLEEVLKVSEEDQSLENSEADVKE SRKDVLP GAEGPASGEGHAG	1358
Db	1228	: : : ~      :    :        :	
Qy	1359	RLSDLGISSYGISETTLEEIFLKVAEE-----SGVDA-ETSDGTLF-----	1267
Qy	1359	NLARCELTSQSASLQSASSVGSARGDEGAGYTDVYGDYRPLF-DNPQDPD--NVSLQEV	1415
Db	1268	:   :         :   : ~: :	
Qy	1416	-----ARRNRRA-FGDKQSCLRPFTEDDAADPNDSIDIPESR	1303
Qy	1416	EAEALS RV-GQGSRKLDGGWLKVRQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMT	1474
Db	1304	:    :  :  ~:     :            :    : :       :	
Qy	1304	ETDLLSGMDGKGSYQVKGWKLTQQQFVALLWKRLLIARRSRKGFFAQIVLPAVFVCI ALV	1363

Qy 1475 VALSVPEIGDLPLVLSPSQYH-NYTQPRGNFIPYANEERREYRLRLSPDASPQQLVSTF 1533  
: | | | | | | | : | | : : : | : : :  
Db 1364 FSLIVPPFGKYPSELELQPWMYNEQYT-----FVSNDAP-----DTGTLELLNAL 1408

Qy 1534 RLPSGVGATCVLKSPANGSLGPTLNLSGESRLAARFFDSMCLESFTQGLPLSNFVPPP 1593  
| | | : : | | :  
Db 1409 TKDPGFGTRCM-----EGNPI----- 1424

Qy 1594 PSPAPSDSPASPDDELQAWNVSPPPTAGPEMWTSAFSLPRLVREPVR-----C 1641  
| | | | | | | : | | : : : : : |  
Db 1425 -----PD-----TPCQAGEEWTAP-VPQTIMDLFQNGNWTMNPSPAC 1463

Qy 1642 TCQAQGTGFS---CPSSVGG-HPPQMRVVTGDILTDTGHNVSLEYLLFTSDRF----- 1690  
| | : | | | | | : | | | | | | | : | | : | | : | :  
Db 1464 QCSSDKIKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKN 1523

Qy 1691 ----RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK----- 1721  
| | | : | | : | : : |  
Db 1524 KIWVNEFRYGGFSLGVSNTQALPPSQEVNDATKQMKKHLKLAKDSSADRFLNSLGRFMTG 1583

Qy 1722 IAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSAS 1781  
: | : : : | | : : : | | : | | | | | | : | | : | | | | | : | |  
Db 1584 LDTRNNVKVWFNNKGWHAISSFLNVINNAAILRANLQKGE-NPSHYGITAFNHPLNLTKQQ 1642

Qy 1782 LS-LDYLLQGTDVVIAIFIIVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLAN 1840  
| | : : | | : : | | | | | | | : | : | | | | : | | : | | : |  
Db 1643 LSEVAPMTTSVDVLVSICVIFAMSFVPASFVFLIQERVSKAKHLQFISGVKPVIIYWLSN 1702

Qy 1841 YVWDMNLNYLVPATCCVIIIFVFDLPAYTSPTNFPVLSLFLLYGWSITPIMYPASFVFEV 1900  
: | | | | | : | | | : | | : | | | | : | | | | | : | | : :  
Db 1703 FVWDMCNVVPATLVIIIFICFQQKSYSSTNLPVLALLLLLYGWSITPLMPASFVFKI 1762

Qy 1901 PSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVVNSYLKSCFLIFPNYNLGHGLME 1960  
| | : | | | : | | | : | | : | | | : | | : | | : | | : :  
Db 1763 PSTAYVVLTSVNLFIGINGSVATFVLELFT-DNKLNNINDILKSVFLIFPHFCLGRGLID 1821

Qy 1961 MAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQR 2020  
| | : : : : | : : : | | | | | | | | : : : : | | | | :  
Db 1822 MVKNQAMADALERFGE-NRFVSPLSWDLVGRNLFAMAVEGVVFFLITVLIQYRFFIRPRP 1880

Qy 2021 MPVSTKPVED-DVDVASERQVRVLRGDADNDMVKIENLTKVYKSRKIGRILAVDRCLGVR 2079  
: : | : | | | | : | | | : | | : | | : | : | : | : | : :  
Db 1881 VNAKLSPLNDEDEDVRRERQRILDGGGQNDILEIKELTKIYRRK---RKPAVDRICVGIP 1937

Qy 2080 PGECFGLLGVNAGAKTSTFKMLTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDAL 2139  
| | | | | | | | : | | | | : | | : | : | : : | : : | | : | :  
Db 1938 PGECFGLLGVNAGAKSSTFKMLTGDTTVTRGDAFLNKNLSILSNIHEVHQNMGYCPQFDAI 1997

Qy 2140 FDELTAREHLQLYTRLRGISWKDEARVVWVALEKLELTKYADKPAGTYSGGNKRKLSTAI 2199  
: | | | : : : | | : | : : | | | | : | | | | | | : :  
Db 1998 TELLTGREHVEFFALLRGVPEKEVGKVGWAIKRLGLVKYGEKYAGNYSGGNKRKLSTAM 2057

Qy 2200 ALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIM 2259  
| | | | : | | | | | | | | | | : | | | | | | | | : | |  
Db 2058 ALIGPPVFLDEPTTGMDPKARRFLWNCALSVVKEGRSVVLTSHSMEECEALCTRM AIM 2117

Qy 2260 VNGRLRCLGSIQHLKNRFGDGYMITVR-TKSSQSVKDVVRFFNRNFP EAM LKERHHTKVQ 2318  
 |||| ||||:||||||| | | | : :| | | | :| :| :|  
 Db 2118 VNGRFRCLGSGVQHLKNRFGDGYTIVVRIAGSNPDLKPVQDFFGLAFPGSV PKEKHRNMLQ 2177

Qy 2319 YQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDNVFNFAKKQSDN 2368  
 ||| | |||:| | : | | ||||| ||||| ||||| |||:  
 Db 2178 YQLPSSLSSLARIFSILSQSKRLHIEDYSVSQTTLQVFNFAKDQSD 2227

#### RESULT 4

##### ABC1\_MOUSE

ID ABC1\_MOUSE STANDARD; PRT; 2261 AA.  
 AC P41233;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE ATP-binding cassette, sub-family A, member 1 (ATP-binding cassette  
 DE transporter 1) (ATP-binding cassette 1) (ABC-1).  
 GN ABCA1 OR ABC1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DBA/2; TISSUE=Macrophage;  
 RX MEDLINE=94375008; PubMed=8088782;  
 RA Luciani M.F., Denizot F., Savary S., Mattei M.-G., Chimini G.;  
 RT "Cloning of two novel ABC transporters mapping on human chromosome  
 RT 9.";  
 RL Genomics 21:150-159(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE=21251004; PubMed=11352567;  
 RA Qiu Y., Cavelier L., Chiu S., Yang X., Rubin E., Cheng J.-F.;  
 RT "Human and mouse ABCA1 comparative sequencing and transgenesis  
 RT studies revealing novel regulatory sequences.";  
 RL Genomics 73:66-76(2001).  
 CC -!- FUNCTION: cAMP-dependent and sulfonylurea-sensitive anion  
 CC transporter. Key gatekeeper influencing intracellular cholesterol  
 CC transport (By similarity).  
 CC -!- TISSUE SPECIFICITY: Widely expressed in adult tissues. Highest  
 CC levels are found in pregnant uterus and uterus.  
 CC -!- DOMAIN: Multifunctional polypeptide with two homologous halves,  
 CC each containing an hydrophobic membrane-anchoring domain and an  
 CC ATP binding cassette (ABC) domain.  
 CC -!- PTM: Phosphorylation on Ser-2054 regulates phospholipid efflux (By  
 CC similarity).  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCA subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

```

CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X75926; CAA53530.1; ALT_INIT.
DR      EMBL; AF287263; AAG39073.1; ALT_INIT.
DR      MGD; MGI:99607; Abcal.
DR      GO; GO:0008203; P:cholesterol metabolism; IDA.
DR      GO; GO:0030301; P:cholesterol transport; IDA.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR003439; ABC_transporter.
DR      Pfam; PF00005; ABC_tran; 2.
DR      ProDom; PD000006; ABC_transporter; 2.
DR      SMART; SM00382; AAA; 2.
DR      PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR      PROSITE; PS50893; ABC_TRANSPORTER_2; 2.
KW      ATP-binding; Glycoprotein; Transmembrane; Transport; Phosphorylation.
FT      TRANSMEM      26      42      POTENTIAL.
FT      TRANSMEM      640     656     POTENTIAL.
FT      TRANSMEM      690     706     POTENTIAL.
FT      TRANSMEM      717     733     POTENTIAL.
FT      TRANSMEM      749     765     POTENTIAL.
FT      TRANSMEM      771     787     POTENTIAL.
FT      TRANSMEM     1041    1057     POTENTIAL.
FT      TRANSMEM     1351    1367     POTENTIAL.
FT      TRANSMEM     1661    1677     POTENTIAL.
FT      TRANSMEM     1708    1724     POTENTIAL.
FT      TRANSMEM     1737    1753     POTENTIAL.
FT      TRANSMEM     1775    1791     POTENTIAL.
FT      TRANSMEM     1854    1870     POTENTIAL.
FT      NP_BIND       933     940     ATP (POTENTIAL).
FT      NP_BIND     1946    1953     ATP (POTENTIAL).
FT      MOD_RES      1042    1042     PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
FT      MOD_RES     2054    2054     PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
FT      CARBOHYD      14      14      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD      98      98      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD     151     151     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD     161     161     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD     196     196     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD     244     244     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD     292     292     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD     337     337     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD     349     349     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD     400     400     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD     478     478     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD     489     489     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD     521     521     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD     820     820     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD    1144    1144     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD    1294    1294     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD    1453    1453     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD    1499    1499     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD    1504    1504     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD    1637    1637     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD    2044    2044     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD    2238    2238     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CONFLICT     1567    1568     MISSING (IN REF. 2).
FT      CONFLICT     2024    2024     MISSING (IN REF. 2).
SQ      SEQUENCE     2261 AA;  254011 MW;  FAE62B21FD1D09F9 CRC64;

```

Query Match 33.1%; Score 4195.5; DB 1; Length 2261;  
Best Local Similarity 39.3%; Pred. No. 9.5e-247;  
Matches 990; Conservative 350; Mismatches 724; Indels 457; Gaps 62;

```
Qy      6 QLQLLLWKNVTLKRRSPWVLAFEIFIPLVLFILLGLRQKKPTISVKEVPFYTAAPL TSA 65
      ||:|||||:| :|| | : || :| ||: :| | | : | : ||
Db      6 QLRLLLWKNLTFRRRQTCQLLLEVAWPLFIFLILISVRLSYPPYEQHECHFNPKA-MPSA 64

Qy     66 GILPVMQSLCPDGQRDEFGFL-----QYANSTVTQLLERLDRVVEEGNLF---DP 112
      | || :| : : | : | : | :| :| :| :| :| :| :| :|
Db     65 GTLPVWQGIICNANNPCFRYPPTGEAPGVVGNFNKSIVSRLEFSDAQRL-----LYSQRDT 120

Qy    113 ARPSLGSELEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQNLSL 172
      : : | || | : | : | : | : | : | : | : | : | : |
Db    121 SIKDMHKVLRMLRQ-----IKHPNSNLKLQDFLVDNETFSGFLQHNLSL 164

Qy    173 PNSTAQALLAARVDPPEVY-----HLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRME 226
      | || :|| | :| : || | : | | : | : | : | : | : |
Db    165 PRSTVDSLQXNVGLQKVFLQGYQLHL-----ASLCNGS-----KLE 201

Qy    227 ELLLAPALLEQLTCTPGSGELGRILTVPESQKGALQGYRDAVCSGQAAARARRFSGLSAE 286
      |: : || | | : : :| : | : | : | : | : | : | : |
Db    202 EII-----QL-----GDAEVSALCGLPRKKL-----DAAERV----- 228

Qy    287 LRNQLDVAK-VSQQLGLDAPNGSDSSPQAPPPRRLQALLGDLLDA----- 330
      || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db    229 LRYNMDILKPVVTKL-----NSTSHLPTQHAEATTVLLDSLGLLAQELFSTKSW 279

Qy    331 ---QKVLQDQDVL SALALLLPQGACTGRTPGPPASGAGGAAN-----GTGAGAVMGP NAT 382
      |:| : | :| : | : | : | : | : | : | : | : | : |
Db    280 DMRQEV MFLTNVNSSSSSTQIYQAVSRIVCGHPEGGGLKIKSLN WYEDNNYKALFGGNNT 339

Qy    383 AEEGAPSAAALATP---DTLQGQCSAFVQ--LWAGLQPILCGNNRTIEPEALRRGNMSSL 437
      |: | : || | : : | : : :| :| :| :| :| :| :|
Db    340 EEDVDTFYDNSTPYCNDLMKNLESSPLSRIIWKALKPLLVG----- 381

Qy    438 GFTSKEQRNLGLLVHMTSNPKILYAPAGSEVDRVILKANETFAFVGNVTHYAQVWLNIS 497
      |||| | :| : :| :| : | :| :| :| :| :| :| :|
Db    382 -----KILYTPDTPATRQVMAEVNKT FQELAVFHDLEGMWHEELS 420

Qy    498 AEIRSFLEQGRLQQHLRWL-----QQYVAELRLHPEAL---NLS 533
      :| :| :| : :| | | : | :| :| :| :| :| :| :| :|
Db    421 PQIWTFMENSQEMDLVRTLLDSRGNDQFWEQKLDGLDWT AQDIMAF LAKNPEDVQSPNGS 480

Qy    534 LDELPPALRQDNFSLPSGMALLOQLDTIDNAACGWIQFMSKVSVDIFKGFPDEESIVNYT 593
      : | : | | : || :|| :|| :|| :| :| :| :| :| :|
Db    481 VYTWREAFNETN-----QAIQTIS-----RFMECVNLNKLEPIPT EURLINKS 523

Qy    594 LNQAYQDNVTVFASVIFQ--TRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTG 651
      : | :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db    524 ME--LLDERKEWAGIVFTGITPDSVELPHHVYKIRMDIDNVERTNKIKDGYWDPGPRAD 581

Qy    652 ---GRFYFLYGFVWIQDMMERAIIDTFVGHDVVEPGSYVQMFPYPCYTRDDFLFVIEHMM 708
      | || :| :| :| :| :| :| :| :| :| :| :| :| :|
Db    582 PFEDMRYVWGGFAYLQDVVEQAIIRVLTGSE-KKTGVYVQQMPYPCYVDDIFLRVMSRSM 640
```

Qy 709 PLCMVISWVYSVAMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAVFITGFGVQLSISVTAL 768  
 || | ::|:||||: |: || ||| |||| |: |||:| : | :||:: : | :| |  
 Db 641 PLFMTLAWIYSVAVIKSIYVEKEARLKETMRIMGLDNGILWFSWFVSSLIPLLVSAGLL 700

Qy 769 TAILKYGQVLMHSHVVIWFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIFYLSYVP 828  
 ||| | :| :| ::::|:|:|: ||: |||:| |:|:| ||:||||||| |:|  
 Db 701 VVILKLGNNLLPYSDPSVVFVFLSVFAMVTILQCFLISTLFSRANLAAACGGIIFYTLYLP 760

Qy 829 YMYVAIREEVVAHDKITAFE-KCIASLMSTTAFGLGSKYFALYEAVGVGIQWHTFSQSPVE 887  
 |: || | | |||:| ||| | :|||:| |:|:| | :|||  
 Db 761 YVLC-----VAWQDYVGFSIKIFASLLSPVAFGFGCEYFALFEEQGIGVQWDLNFESPVE 815

Qy 888 GDDFNLLAVTMLMVDVAVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTEAWE 947  
 | ||| ||:|:| | :||:|||||| || ||:|||||| |||| | | |  
 Db 816 EDGFNLTTAVSMMLFDFTFLYGVMTWYIEAVFPGQYGIPRPWYFPCTKSYWFGE---EIDE 872

Qy 948 WSWPWARTPRLSVMEEDQACAMESRRFEETRGMEEEPHTLPLVVCVDKLTKVYKDDKKLA 1007  
 | | : :| : | | ||||| | | : | |||:| |:|  
 Db 873 KSHPGSSQKGSV-----EIC-----MEEEPHTLRLGVSIQNLVKVYRDGMKVA 915

Qy 1008 LNKLSLNLNLYENQVVSFLGHNGAGKTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNL 1067  
 :: |:| | | |: ||||| ||||| ||||| ||||| |:| | | ||:| | ||:| |  
 Db 916 VDGLALNFYEGQITSFLGHNGAGKTTMSILTGLFPPTSGTAYILGKDIRSEMSSIRQNL 975

Qy 1068 GMCPQHNVLFDRLTVEEHLWFYSRLKSMAQEEIRREMDKMIEDLEL-SNKRHSLVQTL SG 1126  
 |:||||||| |||||:|:|:| |::: :: ||:| |: | :| | |||  
 Db 976 GVCQPQHNVLFDMLTVEEHIWFYARLKLSEKHVKAEMEOMALDVGLPPSKLKSQTL SG 1035

Qy 1127 GMKRKLSVAIAFVGGSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTIILSTHHMDEAD 1186  
 ||:|||||:|||||: : ||||| |||||: || ||:|:|: |||:|||||||  
 Db 1036 GMQRKLSVALAFVGGSKVILDEPTAGVDPYSRRGIWELLLKYRQGRTIILSTHHMDEAD 1095

Qy 1187 LLGDRIAIISHGKLKCCGSPLFLKGTYG DG YRLTLVKRPAEPG----- 1229  
 :||||||| || | ||| | || ||||: |  
 Db 1096 ILGDRIAIISHGKLCCVSSFLKNQLGTGYLTLVKKDVESSLSSCRNSSSTVSCLKKE 1155

Qy 1230 -----GPQEPGLASSPPGRAPLSSCSELQVSQFIRKHAVASCLLVSDTSTELSYILPSEAA 1284  
 : || | | :| ||||: || | ||:| |||  
 Db 1156 DSVSQSSSDAGLSDHESDTLTIDVS--AISNLIRKHVSEARLVEDIGHELTYVLPYEEA 1213

Qy 1285 KKGAFERLFQHLERSLDALHLSSFGLMDTTLEEVLKVSEEDQSLENSEADVKE SRK DVL 1344  
 |:| | | :: | | :||:|: :|||:|:| | | | :| |  
 Db 1214 KEGAFVELFHEIDRLSDLGISSYGISETTLEEIFLKVAEE-----SGVDA-ETSDGTL 1266

Qy 1345 PGAEGPASGEGHAGNLARCSELTQSQASLQSASSVGSARGDEGAGYTDVYGDYRPLF--- 1401  
 | | : | :| :  
 Db 1267 P-----ARRNRA-----FGDKQSCSLHPF 1285

Qy 1402 --DNPQDPD--NVSLOEVEAEALSRV-GQGSRKLDGGWLKVRQFHGLLVKRFHCARRNSK 1456  
 |: ||: :: : | : || : |:| :| | | :|| || || |||: |  
 Db 1286 TEDDAVDPNDSIDIPESRETDLLSGMDGKGSYQLKGWKL TQQFVALLWKRLLIARRSRK 1345

Qy 1457 ALFSQILLPAFFVCVAMTVALSVPEIGDLPLVLSPSQYH-NYTQPRGNFIPYANEERRE 1515  
 |:|:| | ||:|: :| || | | | | |: || : : : |  
 Db 1346 GFFAQIVLPAVFVCIALVFSLIVPPFGKYPSELELQPMYNEQYT-----FVSNDAPE 1397

Qy 1516 YRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFFDSM 1575



Db	1398	-----DMGTQELLNALT KDPGFGTRCMEGNPIP-----DTP	1428
Qy	1576	CLESFTQGLPLSNFVPPPPSPAPSDSPASPDDELQAWNVS LPPPTAGPEM-----WTSAPS	1630
Db	1429	CL-----AGEED---WTISPVPQSIVDLFQNGNWTMKNP	1459
Qy	1631	LPRLVREPVRCTCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILTITGHNVS EYLLFT	1686
Db	1460	SP-----ACQCSSDKIKMLPVCPPGAGGLPPPQRKQKTADILQNL TGRNISDYL VKT	1512
Qy	1687	SDRF-----RLHRYGAITFG-----NVLKSIPASFGTRAPP	1717
Db	1513	YVQIIAKSLKNKIWVNEFRYGGFSLGVSNSQALPPSHEVND AIKQMKLLKLT KDTSADR	1572
Qy	1718	MVRKIA-----VRRAAQVFYNNKGYHSMPTYLNSLN NAILRANLPKSKGNPAAYGITV	1770
Db	1573	FLSSLGRFMAGLDTKNNVKVWFNNKGWHAISSFLN VINNAILRANLQKGE-NPSQYGITA	1631
Qy	1771	TNHPMNKTSASLS-LDYLLQGTDVVIAIFIIVAMS FVPASFVVFLVAEKSTKAKHLQFVS	1829
Db	1632	FNHPLNLTKQQLSEVALMTTSVDVLVSICVIFAMSFVPASFVV FLIQERVSKAKHLQFIS	1691
Qy	1830	GCPNPIIYWLANYVWDMNLNYLVPATCCVII LFVFDLPAYTSPTNFP AVLSLFLLYGWSITP	1889
Db	1692	GVPKVIYWLSNFWDMCNVVPATLVII FICFQQKSYVSSTNLPVLALLLLLYGWSITP	1751
Qy	1890	IMYPASFWFEVPSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSY LKSCFLIF	1949
Db	1752	LMYPASFVKIPSTAYVVLTSVNLFIGINGSVATFVLELFTNNK-LNDINDILKSVFLIF	1810
Qy	1950	PNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVG FLLTIM	2009
Db	1811	PHFCLGRGLIDMVKNQAMADALERFGE-NRFVSPLSWDLVGRNLFAMAVEGVVFFLITVL	1869
Qy	2010	CQYNFLRRPQRMVPVSTKPVED-DVDVASERQVR LRGDADNDMVKIENLTKVYKSRKIGRI	2068
Db	1870	IQYRFFIRPRPVKAKLPPLNDEDEDVRRERQRILDGGGQNDILEIKELTKIYRRK---RK	1926
Qy	2069	LAVDRLCIGVRPGECFGLLG VNGAGKTSTFKMLTGDESTTGGEAFVNGHSVLKELLQVQQ	2128
Db	1927	PAVDRICIGIPPGEFCFGLLG VNGAGKSTTFKMLTGDTPVTRGDAFLNKN S ILSNIHEVHQ	1986
Qy	2129	SLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVV KWALEKLELTKYADKPAGTYS	2188
Db	1987	NMGYCPQFDAITELLTGREHVEFFALLRGVPEKEVGKFG EWAIRKLG LVKYGEKYASNYS	2046
Qy	2189	GGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWN LILD LIKTGRSVVLTSHSMEE	2248
Db	2047	GGNKRKLSTAMALIGPPVVFLDEPTTGMDPKARRFLWNCAL SIVKEGRSVVLTSHSMEE	2106
Qy	2249	CEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVR-TKSSQSVKDVVRFFN RN FPEA	2307
Db	2107	CEALCTRM AIMVNGRFRCLG SVQHLKNRFGDGYTIVVRIAGSNPDLKP VQE FGLAFPGS	2166
Qy	2308	MLKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQT TLDNVFVNFAKKQSD	2367

Db 2167 VLKEKHRNMLQYQLPSSLSSLARIFSILSQSKRLHIEDYSVSQTTLDQVFVNFAKDQSD 2226

Qy 2368 N 2368

:

Db 2227 D 2227

RESULT 5

ABCR\_HUMAN

ID ABCR\_HUMAN STANDARD; PRT; 2273 AA.

AC P78363; O15112; O60438; O60915;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Retinal-specific ATP-binding cassette transporter (RIM ABC transporter) (RIM protein) (RMP) (Stargardt disease protein).

GN ABCA4 OR ABCR.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., VARIANTS STGD, AND VARIANTS HIS-846 AND GLN-943.

RX MEDLINE=97207641; PubMed=9054934;

RA Allikmets R., Singh N., Sun H., Shroyer N.F., Hutchinson A.,

RA Chidambaram A., Gerrard B., Baird L., Stauffer D., Peiffer A.,

RA Rattner A., Smallwood P.M., Li Y., Anderson K.L., Lewis R.A.,

RA Nathans J., Leppert M., Dean M., Lupski J.R.;

RT "A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is

RT mutated in recessive Stargardt macular dystrophy.";

RL Nat. Genet. 15:236-246(1997).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=97345663; PubMed=9202155;

RA Azarian S.M., Travis G.H.;

RT "The photoreceptor rim protein is an ABC transporter encoded by the

RT gene for recessive Stargardt's disease (ABCR).";

RL FEBS Lett. 409:247-252(1997).

RN [3]

RP SEQUENCE FROM N.A., AND VARIANTS STGD TRP-18 AND CYS-212.

RX MEDLINE=98163759; PubMed=9503029;

RA Gerber S., Rozet J.-M., van de Pol T.J.R., Hoyng C.B., Munnich A.,

RA Blankenagel A., Kaplan J., Cremers F.P.M.;

RT "Complete exon-intron structure of the retina-specific ATP binding

RT transporter gene (ABCR) allows the identification of novel mutations

RT underlying Stargardt disease.";

RL Genomics 48:139-142(1998).

RN [4]

RP SEQUENCE FROM N.A., AND VARIANTS STGD.

RX MEDLINE=98141123; PubMed=9490294;

RA Nasonkin I., Illing M., Koehler M.R., Schmid M., Molday R.S.,

RA Weber B.H.F.;

RT "Mapping of the rod photoreceptor ABC transporter (ABCR) to 1p21-p22.1

RT and identification of novel mutations in Stargardt's disease.";

RL Hum. Genet. 102:21-26(1998).

RN [5]

RP CHARACTERIZATION.

RX MEDLINE=99175213; PubMed=10075733;  
 RA Sun H., Molday R.S., Nathans J.;  
 RT "Retinal stimulates ATP hydrolysis by purified and reconstituted ABCR,  
 RT the photoreceptor-specific ATP-binding cassette transporter  
 RT responsible for Stargardt disease.";  
 RL J. Biol. Chem. 274:8269-8281(1999).  
 RN [6]  
 RP DISEASE.  
 RX MEDLINE=98133912; PubMed=9466990;  
 RA Cremers F.P.M., van de Pol D.J.R., van Driel M.A., den Hollander A.I.,  
 RA van Haren F.J.J., Knoers N.V.A.M., Tijmes N., Bergen A.A.B.,  
 RA Rohrschneider K., Blankenagel A., Pinckers A.J.L.G., Deutman A.F.,  
 RA Hoyng C.B.;  
 RT "Autosomal recessive retinitis pigmentosa and cone-rod dystrophy  
 RT caused by splice site mutations in the Stargardt's disease gene  
 RT ABCR.";  
 RL Hum. Mol. Genet. 7:355-362(1998).  
 RN [7]  
 RP VARIANTS ARMD2, AND VARIANTS.  
 RX MEDLINE=97442530; PubMed=9295268;  
 RA Allikmets R., Shroyer N.F., Singh N., Seddon J.M., Lewis R.A.,  
 RA Bernstein P.S., Peiffer A., Zabriskie N.A., Li Y., Hutchinson A.,  
 RA Dean M., Lupski J.R., Leppert M.;  
 RT "Mutation of the Stargardt disease gene (ABCR) in age-related macular  
 RT degeneration.";  
 RL Science 277:1805-1807(1997).  
 RN [8]  
 RP VARIANTS STGD TRP-18; CYS-212; HIS-636; MET-1019; VAL-1038; CYS-1108;  
 RP TRP-1640; SER-1977 AND HIS-2107, AND VARIANTS FFM PRO-11; PRO-541;  
 RP VAL-1038; GLU-1091; CYS-1508; PHE-1970 AND ARG-1971.  
 RX MEDLINE=98454319; PubMed=9781034;  
 RA Rozet J.-M., Gerber S., Souied E., Perrault I., Chatelin S., Ghazi I.,  
 RA Leowski C., Dufier J.-L., Munnich A., Kaplan J.;  
 RT "Spectrum of ABCR gene mutations in autosomal recessive macular  
 RT dystrophies.";  
 RL Eur. J. Hum. Genet. 6:291-295(1998).  
 RN [9]  
 RP VARIANTS STGD.  
 RX MEDLINE=99138655; PubMed=9973280;  
 RA Lewis R.A., Shroyer N.F., Singh N., Allikmets R., Hutchinson A.,  
 RA Li Y., Lupski J.R., Leppert M., Dean M.;  
 RT "Genotype/phenotype analysis of a photoreceptor-specific ATP-binding  
 RT cassette transporter gene, ABCR, in Stargardt disease.";  
 RL Am. J. Hum. Genet. 64:422-434(1999).  
 RN [10]  
 RP VARIANTS STGD, AND VARIANTS.  
 RX MEDLINE=99192348; PubMed=10090887;  
 RA Maugeri A., van Driel M.A., van de Pol D.J.R., Klevering B.J.,  
 RA van Haren F.J.J., Tijmes N., Bergen A.A.B., Rohrschneider K.,  
 RA Blankenagel A., Pinckers A.J.L.G., Dahl N., Brunner H.G.,  
 RA Deutman A.F., Hoyng C.B., Cremers F.P.M.;  
 RT "The 2588G-->C mutation in the ABCR gene is a mild frequent founder  
 RT mutation in the western European population and allows the  
 RT classification of ABCR Mutations in patients with Stargardt disease.";  
 RL Am. J. Hum. Genet. 64:1024-1035(1999).  
 RN [11]  
 RP VARIANT STGD TYR-54, AND VARIANT ALA-863.

RX MEDLINE=20077755; PubMed=10612508;  
 RA Zhang K., Garibaldi D.C., Kniazeva M., Albini T., Chiang M.F.,  
 RA Kerrigan M., Sunness J.S., Han M., Allikmets R.;  
 RT "A novel mutation in the ABCR gene in four patients with autosomal  
 RT recessive Stargardt disease.";  
 RL Am. J. Ophthalmol. 128:720-724(1999).  
 RN [12]  
 RP VARIANTS STGD VAL-60; ARG-206; ASN-300; PRO-541; ALA-849; PRO-974;  
 RP VAL-1038; CYS-1108; LEU-1408; ARG-1488; ASP-1652; PRO-1729; GLU-1961;  
 RP TRP-2038; TRP-2077; HIS-2107; ARG-2128 AND TYR-2150.  
 RX MEDLINE=99221420; PubMed=10206579;  
 RA Fishman G.A., Stone E.M., Grover S., Derlacki D.J., Haines H.L.,  
 RA Hockey R.R.;  
 RT "Variation of clinical expression in patients with Stargardt dystrophy  
 RT and sequence variations in the ABCR gene.";  
 RL Arch. Ophthalmol. 117:504-510(1999).  
 RN [13]  
 RP VARIANTS GLU-1961 AND ASN-2177.  
 RX MEDLINE=20349288; PubMed=10880298;  
 RA Allikmets R., Tammur J., Hutchinson A., Lewis R.A., Shroyer N.F.,  
 RA Dalakishvili K., Lupski J.R., Steiner K., Pauleikhoff D., Holz F.G.,  
 RA Weber B.H.F., Dean M., Atkinson A., Gail M.H., Bernstein P.S.,  
 RA Singh N., Peiffer A., Zabriskie N.A., Leppert M., Seddon J.M.,  
 RA Zhang K., Sunness J.S., Udar N.S., Yelchits S., Silva-Garcia R.,  
 RA Small K.W., Simonelli F., Testa F., D'Urso M., Brancato R.,  
 RA Rinaldi E., Ingvasst S., Taube A., Wadelius C., Souied E., Ducroq D.,  
 RA Kaplan J., Assink J.J.M., ten Brink J.B., de Jong P.T.V.M.,  
 RA Bergen A.A.B., Maugeri A., van Driel M.A., Hoyng C.B., Cremers F.P.M.,  
 RA Paloma E., Coco R., Balcells S., Gonzalez-Duarte R., Kermani S.,  
 RA Stanga P., Bhattacharya S.S., Bird A.C.;  
 RT "Further evidence for an association of ABCR alleles with age-related  
 RT macular degeneration.";  
 RL Am. J. Hum. Genet. 67:487-491(2000).  
 RN [14]  
 RP VARIANTS STGD GLU-60; THR-60; GLU-65; LEU-68; ARG-72; CYS-212;  
 RP SER-230; SER-247; VAL-328; LYS-471; PRO-541; GLN-572; ARG-607;  
 RP LYS-635; CYS-653; TYR-764; ARG-765; ALA-901; ILE-959; LYS-1036;  
 RP VAL-1038; PRO-1063; ASP-1087; CYS-1097; CYS-1108; LEU-1380; LYS-1399;  
 RP PRO-1430; VAL-1440; HIS-1443; LEU-1486; TYR-1488; MET-1537; PRO-1689;  
 RP LEU-1705; THR-1733; ARG-1748; PRO-1763; LYS-1885; HIS-1898; GLU-1961;  
 RP ARG-1975; SER-1977; GLY-2077; TRP-2077 AND VAL-2241, AND VARIANTS  
 RP GLN-152; HIS-212; ARG-423; ILE-552; ARG-914; GLN-943; THR-1562;  
 RP ILE-1868; MET-1921; LEU-1948; PHE-1970; ALA-2059; ASN-2177 AND  
 RP VAL-2216.  
 RX MEDLINE=20442027; PubMed=10958763;  
 RA Rivera A., White K., Stoehr H., Steiner K., Hemmrich N., Grimm T.,  
 RA Jurklies B., Lorenz B., Scholl H.P.N., Apfelstedt-Sylla E.,  
 RA Weber B.H.F.;  
 RT "A comprehensive survey of sequence variation in the ABCA4 (ABCR) gene  
 RT in Stargardt disease and age-related macular degeneration.";  
 RL Am. J. Hum. Genet. 67:800-813(2000).  
 RN [15]  
 RP VARIANTS CORD3 GLU-65; CYS-212; PRO-541; ALA-863; GLY-863 DEL;  
 RP VAL-1038; LYS-1122; TYR-1490 AND ASP-1598.  
 RX MEDLINE=20442040; PubMed=10958761;  
 RA Maugeri A., Klevering B.J., Rohrschneider K., Blankenagel A.,  
 RA Brunner H.G., Deutman A.F., Hoyng C.B., Cremers F.P.M.;

RT "Mutations in the ABCA4 (ABCR) gene are the major cause of autosomal  
 RT recessive cone-rod dystrophy.";  
 RL Am. J. Hum. Genet. 67:960-966(2000).  
 RN [16]  
 RP VARIANTS STGD ASP-340; GLN-572; ALA-863; SER-965; VAL-1038; ALA-1780  
 RP AND HIS-1898, AND VARIANT GLN-943.  
 RX MEDLINE=20208356; PubMed=10746567;  
 RA Shroyer N.F., Lewis R.A., Lupski J.R.;  
 RT "Complex inheritance of ABCR mutations in Stargardt disease: linkage  
 RT disequilibrium, complex alleles, and pseudodominance.";  
 RL Hum. Genet. 106:244-248(2000).  
 RN [17]  
 RP VARIANTS STGD.  
 RX MEDLINE=20098082; PubMed=10634594;  
 RA Papaioannou M., Ocaka L., Bessant D., Lois N., Bird A.C., Payne A.,  
 RA Bhattacharya S.S.;  
 RT "An analysis of ABCR mutations in British patients with recessive  
 RT retinal dystrophies.";  
 RL Invest. Ophthalmol. Vis. Sci. 41:16-19(2000).  
 RN [18]  
 RP VARIANTS STGD CYS-212; ASP-767; ILE-897; VAL-1038; LYS-1087; LYS-1399;  
 RP GLN-1640 AND GLU-1961, AND VARIANT HIS-212.  
 RX MEDLINE=20174852; PubMed=10711710;  
 RA Simonelli F., Testa F., de Crecchio G., Rinaldi E., Hutchinson A.,  
 RA Atkinson A., Dean M., D'Urso M., Allikmets R.;  
 RT "New ABCR mutations and clinical phenotype in Italian patients with  
 RT Stargardt disease.";  
 RL Invest. Ophthalmol. Vis. Sci. 41:892-897(2000).  
 RN [19]  
 RP CHARACTERIZATION OF VARIANTS, AND MUTAGENESIS OF GLY-966; LYS-969;  
 RP GLY-1975 AND LYS-1978.  
 RX MEDLINE=20472331; PubMed=11017087;  
 RA Sun H., Smallwood P.M., Nathans J.;  
 RT "Biochemical defects in ABCR protein variants associated with human  
 RT retinopathies.";  
 RL Nat. Genet. 26:242-246(2000).  
 RN [20]  
 RP VARIANT STGD ASN-972, AND VARIANTS GLN-943; ILE-1868 AND LEU-1948.

Query Match 30.6%; Score 3875.5; DB 1; Length 2273;  
 Best Local Similarity 35.7%; Pred. No. 3e-227;  
 Matches 910; Conservative 403; Mismatches 763; Indels 471; Gaps 58;

Qy 1 MGFLHQLQLLLWKNVTLKRRSPWVLAFEIFIPLVLFFILLGLRQKKPTISVKEVPFYTAA 60  
 |||: |:||||| ||::| |: || |:|: || | | | |  
 Db 1 MGFEVRQIQLLLWKNWTLRKRQKIRFVVELVWPLSLFLVLIWLRNANPLYSHHECHFPNKA 60  
 Qy 61 PLTSAGILPVMQSLCPDGQRDEF-----GFLQYANSTVTQLLERLDRVVEEGLNLF 111  
 : |||: || :| : : | : |:: | |: | :| :  
 Db 61 -MPSAGMLPWLQGIFCNVNNPCFQSPTPGESPGIVSNYNNSI---LARVYRDFQELLMNA 116  
 Qy 112 PARPSLG---SELEALRQHLEALSAGPGTSGSHLDRSTVSSFSLDSVARNPQELWRFLTQ 168  
 | || :|| | |:: | :| :| : : : : | || :  
 Db 117 PESQHLGRIWTELHILSQFMDTLR-----THPERIAGRIRIRDILKDEETLTFLIK 169  
 Qy 169 NLSLPNSTAQALLAARVDPPEVYHLLFGPSSALDSQSGLHKGQEPWSRLGGNPLFRMEEL 228  
 |: | :| |: ::| | : | | ::::

Db	170	NIGLSDSVVYLLINSQVRPEQFAH-----GVPDLALKDI	203
Qy	229	LLAPALLEQLTCTPGSGELGRILTVPESQKQALQGYRDAVCSGQAAARARRFSGLSAELR	288
		:    : :   : :    :  :	
Db	204	ACSEALLERF-----IIFSQRGAKTVRYALCSLSQGT----LQWIEDTLY	245
Qy	289	NQLDVAKVSQQLGLDAPNGSDSSPQAPPPRRLQALLGDLLDAQKVLQDV-----DVLSA	342
		:   : :            :   : : : : :  :	
Db	246	ANVDFFKLFRVL----PTLLDSRSQGINLRSWGGILSDM--SPRIQEFIHREPSMQDLLWV	299
Qy	343	LALLLPQGACTGRTPGPPASGAGGAANGTGAGAVMGPNATAEEGAPSAAALATPDTLQGG	402
		:   :	
Db	300	TRPLMQNGG-----PET----	311
Qy	403	CSAFVQLWAGLQPILCGNNTIEPEALRRGNMSSLGFTSKEQRN----LGL-----	449
		:    :                :	
Db	312	---FTKLMGILSDLLCG----YPEG---GGSRVLSFNWYEDNNYKAFLGIDSTRKDPIY	360
Qy	450	-----LVHLMTSNP-----KILYAPAGSEVDRVILKANETFAF	482
		: :            :	
Db	361	SYDRRTTSFCNALIQSLESNPLTKIAWRAAKPLLMGKILYTPDSPAARRILKNANSTFEE	420
Qy	483	VGNVTHYAQVWLNISAEIRSFLEQGRLLQOHLR-----WLQQYVAELRLHPEALNL	532
		: :  :   : :    : :  : :   :   :	
Db	421	LEHVRKLVKAWEEVGPQIWYFFDNSTQMMNIRDTLGNPTVKDFLNRQLGEEGITAEAILN	480
Qy	533	SLDELPPALRQD---NFSLPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFPEDESI	589
		:  :      : : :     : : :   : :    :	
Db	481	FLYKGPRESQADDMANFWRDIFNITDRTLRLVN-----QYLECLVLDKFESYNDETQL	534
Qy	590	VNYTLNQAYQDNVTVFASFVIFQTRK--DGSLLPHVHYKIRQNSSFTEKTNEIRRAYWRPG	647
		: :  : :   :            :    : :	
Db	535	TQRALS-LLEENM-FWAGVVFPDMPWTSSLPHVKYKIRMDIDVVEKTNKIKDRYWDG	592
Qy	648	PNTGGRFYFLY---GFVWIQDMMERAIIDTFVGHDVVEPGSYVQMFPYPCYTRDDFLFVI	704
		:  : :   :   :   :     :    : :	
Db	593	PRADPVEDFRYIWGGFAYLQDMVEQGITRSQVQAE-APVGIYLQMPYPFCVDDSFMIIL	651
Qy	705	EHMMPLCMVISWVYSVAMTIQHIVAEKEHRLKEVMKTMGLNNAVHWVAFITGFVQLSIS	764
		:   : : : : :            :   :        :   : :	
Db	652	NRCFPIFMVLAWIYSVSMTVKSIVLEKELRLKETLNQGVSSNAVIWCTWFLDSDFSIMSMS	711
Qy	765	VTALTAILKYQVLMHSHVVIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFL	824
		:    : : :  :   :    : :        :  :     :   :	
Db	712	IFLLTIFIMHGRIILHYSDFILFLFLLAFSTATIMLCFLLSTFFSKASLAAACSGVIYFT	771
Qy	825	SYVPYMYVAIREEVAHDKITAFEKCIASLMSTTAFGLGSKYFALYEAVAGVGIQWHTFSQS	884
		: :~ :  :       :      :  :   : :	
Db	772	LYLPHILCFAWQ----DRMTAELKKAVALSLSPVAFGFGTEYLVRFEEOGLGLQWSNIGNS	827
Qy	885	PVEGDDFNLLLAVTMLMVDVVYGILTWYIEAVHPGMYGLPRPWYFPLQKSYWLGSGRTE	944
		: :   :   : :     :    :               : :	
Db	828	PTGEGDEFSLLSMQMMLLDAAVYGLLAWYLDQVFPGDYGTPLPWYFLLQESYWLGG----	883
Qy	945	AWESWPPWARTPRLSVMEEDQACA-MESRRFEETRGMEEE---PTHLPV-----	990
		:  :      :  :        :	
Db	884	-----EGCSTREERALEKTEPLTEETEDPEHPEGIHDSFFEREHP	923

Qy 991 -----VCVDKLTKVYKDDKKLALNKLNLNLYENQVVSFLGHNGAGKTTTMSILTGLFPPT 1045  
 ||| | |::: : |:::|::: ||||: :|||||||:||||| |||  
 Db 924 GWVPGVCVKNLVKIFEPGRPAVDRNLNITFYENQITAFGLGHNGAGKTTTSLILTGLLPPT 983  
 Qy 1046 SGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSRLKSMAQEEIRREMD 1105  
 ||: : | || | :| :|::| |||||:| | || ||: ||:| | :|| : ||:  
 Db 984 SGTVLVGGRDIETSLDAVRQSLGMCPQHNLFHHLTVAEHMLFYAQLKGKSQEEAQLME 1043  
 Qy 1106 KMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAFVGGSRAILDEPTAGVDPYARRAIWDL 1165  
 |::| | :||: | ||||:||||||| : : |||||:|||||:|:| |||  
 Db 1044 AMLEDGLHHRNEEAQDLSSGGMQRKLSVAIAFVGDAKVVILDEPTSGVDPYSRRSIWDL 1103  
 Qy 1166 ILKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPLFLKGTYGDDYRLTLVKRP 1225  
 :|||: ||||:| ||||| |||||: | :| | | :||| | :| | ||||:|  
 Db 1104 LLKYRSGRTIIMSTHHMDEADLLGDRIAIIAQGRLYCSGTPLFLKNCFGTGLYTLVLRK- 1162  
 Qy 1226 AEPGGPQEPGLASSPPGRAPLSSCSEL-----QVSQFIRKH 1261  
 : | | ||| : : : |  
 Db 1163 -----MKNIQSQRKSGEGTSCSSKGFSTTCPAHVDDLTPEQVLDGDVNEIMDVVLHH 1215  
 Qy 1262 VASCLLVSDTSTELSYILPSEAAKKAFAERLFQHLERSLDALHLSSFGLMDTTLEEVFLK 1321  
 | || || :||:| | : ||: || :| | ||||: || |||:| |||  
 Db 1216 VPEAKLVEICIGQELIFLLPNKNFKHRAYASLFRELEETLADLGLSSFGISDTPLEEIFLK 1275  
 Qy 1322 VSEEDQSLENSEADVKESSRKDVLPAGEGPASGEGHAGNLARCSELTQSQASLQSASSVGS 1381  
 |::: | : : |::| | | |  
 Db 1276 VTEDSDSGPLFAGGAQQKRENVNP--RHPCLGPE----- 1306  
 Qy 1382 ARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEEAEALS RVGQ-----GSRKLDGG 1433  
 | || | ||| : | | | || | : |  
 Db 1307 ---REKAGQT-----PQDSNVCS---PGAPAAHPEGQPPPECEPCGPQLNTGT 1348  
 Qy 1434 WLKVRQFHGLLVKRFHCARRNSKALFSQILLPAFFVFCVAMTVALSVPEIGDLPLVLSPS 1493  
 | : : ||||| | : | :||:| | | :| : : : | | : | | | |  
 Db 1349 QLVQLHVQALLVKRFQHTIRSHKDFLAQIVLPATFVFLALMLSIVIPPFGEYPALTLHPW 1408  
 Qy 1494 QYHNYTQPRGNFIPYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCVLKSPANGSL 1553  
 | :|| : : | | :| |  
 Db 1409 IY-----GQQYTFFSMDEPGSEQFT-----VLADVLLNKPG--- 1439  
 Qy 1554 GPTLNLSSGESRLLAARFFDSMCL-ESFTQGLPLSNFVPPPPSPAPSDSPASPDDELQAW 1612  
 | : || | : | | | |||  
 Db 1440 -----FGNRCLKEGWLPEYPCGN---STPWKTPSVSP----- 1468  
 Qy 1613 NVSLPPTAGPEMWTSAPSLRLVREPVRCTCSAQGTGF---SCPSSVGGHPPQMRVV-TG 1668  
 : : : || | || : || || | | :  
 Db 1469 ---NITQLFQKQKWTQVNPSP-----SCRCSTREKLTMLPECPEGAGGLPPPQRTQRTST 1519  
 Qy 1669 DILTDITGHNVSEYLLFT-----SDRFR LH--RYGAITFGNVLKSIPAS----- 1710  
 :|| | :| | :||:| | :| : ||| | : | | :| :  
 Db 1520 EILQDLTDRNISDFLVKTYPALIRSSLKSKFWVNEQRYGGISIGGKLPVVPITGEALVGF 1579  
 Qy 1711 -----FGTRAPPMVRKIAVRRAAQVFYNNKGYHSMPTYLNSLNNAIL 1752  
 | : : :|||:|:| :|| :|||  
 Db 1580 LSDLGRIMNVSGGPITREASKEIPDFLKHLETEDNIKVWFNNKGWHALVSFLNVAHNAIL 1639





OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Thyroid carcinoma;  
 RX MEDLINE=96326608; PubMed=8706931;  
 RA Klugbauer N., Hofmann F.;  
 RT "Primary structure of a novel ABC transporter with a chromosomal  
 RT localization on the band encoding the multidrug resistance-associated  
 RT protein.";  
 RL FEBS Lett. 391:61-65(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97179225; PubMed=9027511;  
 RA Connors T.D., van Raay T.J., Petry L.R., Klinger K.W., Landes G.M.,  
 RA Burn T.C.;  
 RT "The cloning of a human ABC gene (ABC3) mapping to chromosome  
 RT 16p13.3.";  
 RL Genomics 39:231-234(1997).  
 CC -!- FUNCTION: May be a transporter, its natural substrate has not been  
 CC found yet (By similarity). May act as an efflux pump for  
 CC chemotherapeutics drugs.  
 CC -!- TISSUE SPECIFICITY: Highly expressed in lung, followed by brain,  
 CC pancreas, skeletal muscle and heart. Weakly expressed in placenta,  
 CC kidney and liver. Also expressed in medullary thyroid carcinoma  
 CC cells (MTC) and in C-cell carcinoma.  
 CC -!- DOMAIN: Multifunctional polypeptide with two homologous halves,  
 CC each containing an hydrophobic membrane-anchoring domain and an  
 CC ATP binding cassette (ABC) domain (By similarity).  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCA subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U78735; AAC50967.1; -.  
 DR EMBL; X97187; CAA65825.1; -.  
 DR PIR; A59188; A59188.  
 DR PIR; S71363; S71363.  
 DR Genew; HGNC:33; ABCA3.  
 DR MIM; 601615; -.  
 DR GO; GO:0016021; C:integral to membrane; TAS.  
 DR GO; GO:0005624; C:membrane fraction; TAS.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; TAS.  
 DR GO; GO:0005215; F:transporter activity; TAS.  
 DR GO; GO:0009315; P:drug resistance; TAS.  
 DR GO; GO:0006810; P:transport; TAS.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 2.  
 DR ProDom; PD000006; ABC\_transporter; 2.  
 DR SMART; SM00382; AAA; 2.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.

DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
 KW ATP-binding; Transport; Transmembrane.  
 FT TRANSMEM 22 42 POTENTIAL.  
 FT TRANSMEM 249 269 POTENTIAL.  
 FT TRANSMEM 307 327 POTENTIAL.  
 FT TRANSMEM 344 364 POTENTIAL.  
 FT TRANSMEM 373 393 POTENTIAL.  
 FT TRANSMEM 405 425 POTENTIAL.  
 FT TRANSMEM 447 467 POTENTIAL.  
 FT TRANSMEM 925 945 POTENTIAL.  
 FT TRANSMEM 1100 1120 POTENTIAL.  
 FT TRANSMEM 1144 1164 POTENTIAL.  
 FT TRANSMEM 1183 1203 POTENTIAL.  
 FT TRANSMEM 1213 1233 POTENTIAL.  
 FT TRANSMEM 1245 1265 POTENTIAL.  
 FT TRANSMEM 1306 1326 POTENTIAL.  
 FT NP\_BIND 566 573 ATP (POTENTIAL).  
 FT NP\_BIND 1416 1423 ATP (POTENTIAL).  
 FT CONFLICT 36 36 P -> S (IN REF. 2).  
 FT CONFLICT 196 196 L -> P (IN REF. 2).  
 SQ SEQUENCE 1704 AA; 191387 MW; AF0098DAF7A04F5F CRC64;

Query Match 20.7%; Score 2622; DB 1; Length 1704;  
 Best Local Similarity 34.0%; Pred. No. 4.3e-151;  
 Matches 638; Conservative 317; Mismatches 556; Indels 364; Gaps 45;

Qy 581 KGFPDEESIVNYTLNQAYQDNV--TVFASVIFQ---TRKDGSLPPHVHYKIR----- 627  
 :||| |: |: || |:|:|: || ||| :|  
 Db 108 RGFPSEKDFEDY----IRYDNCSSSVLAAVVFEHPFNHSKEPLPLAVKYHLRFSYTRRNY 163  
 Qy 628 ---QNSSFTEK-----TNEIRRAYWRPG-----PNTGGRFYFLYGFVWIQDMMERAI 672  
 | || | : : || |: | | ||: :| ::||:  
 Db 164 MWTQTGSFFLKETEGWHTTSLFPLFPNPGPRELTSPDGGEPPGYIREGFLAVQHAVDRAIM 223  
 Qy 673 DTFVGHDVVEPGSY-----VQMFYPCYTRDDFLFVIEHMMPLCMVISWVYSVAMTIQH 726  
 : | : : ||| : | || |: :| ::|: | : :  
 Db 224 EYHA--DAATRQLFQRLTVTIKRFYPYPPFIADPFLVAIQYQLPLLLLLSFTYTALTIA 281  
 Qy 727 IVAEKEHRLKEVMKTMGLNNAVHWVAVFITGFVQLSISVTALTAL-----KYGQVLMHS 781  
 :| ||| |||| |: |||: :| ||| :| | |: :| : |||  
 Db 282 VVQEKERRLKEYMRMMGLSSWLHWSAWFLFLFLLIAASFMTLLFCVKVKNVAVLSRS 341  
 Qy 782 HVVIIWLFLAVYAVATIMFCFLVSVLYSKAKLASACGGIIYFLSYVPYMYVAIREEVAHD 841  
 :: || :|:| | |:| :|| :|:| || :| :|:| :|| | ::  
 Db 342 DPSLVLAFLLCFAISTISFSFMVSTFFSKANMAAAGGFLYFFTYIPYFFVAPR----YN 397  
 Qy 842 KITAFEKCIASLMSTTAFGLGSKYFALYEVAGVGIQWHTFSQSPVE-GDDFNLLLAVTML 900  
 :| :| : |:| | :|: :| |:| | || ||| : ||  
 Db 398 WMTLSQKLCSCLLSNVAMAMGAQLIGKFEAKGMGIQWRDL-LSPVNVDDDFCFGQVLGML 456  
 Qy 901 MVDAVVYGILTWYIEAVHPPGMYGLPRPWYFPLQKSYWLGSGRTEAWESWPPWARTPRLSV 960  
 ::|:|:|:|:|:| | | :|:|:| | : ||| | | :  
 Db 457 LLDSVLYGLVTWYMEAVFPGQFGVPQWPYFFIMPYSWCGKPRVAVAGK----- 503  
 Qy 961 MEEDQACAMESRRFEETRGMEEPTHLPVVCVDKLTQVYK--DDKKLALNKLNLNLYEN 1018  
 ||: : :| | | | : : |:|: : : |: | ||||  
 Db 504 -EEEDSDPEKALRNEY---FEAEPEDLVAGIKIKHLSKVFRVGNKDRAAVRDLNLNLYEG 559

Qy	1019	QVVSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLF	1078
Db	560	QITVLLGHNGAGKTTTMSLTGLFPPTSGRAYISGYEISQDMVQIRKSLGLCPQHDILFD	619
Qy	1079	RLTVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSSGGMKRKLSVAIAF	1138
Db	620	NLTVAEHLFYAQLKGLSRQKCPEEVKQMLHIIGLEDKWNRSRSLSGGMRRKLSIGIAL	679
Qy	1139	VGGSRAIILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEADLLGDRIAIISHG	1198
Db	680	IAGSKVLILDEPTSGMDAISRRAIWDLIQRQKSDRTIVLTTHFMDEADLLGDRIAIMAKG	739
Qy	1199	KLKCCGSPLFLKGTYG DG YRLTLVKRPAEPGGPQEPGLASSPFGRAPLSSCSELQVSQFI	1258
Db	740	ELQCCGSSLFLKQKYGAGYHMTLVKEP-----HCNPEDISQLV	777
Qy	1259	RKHVASCLLVSDTSTELSYILPSEAAKKGAFERLFQHLERSLDALHLSSFGLMDTTLEEV	1318
Db	778	HHHVPNATLESSAGAELSFILPRESTHR--FEGLFAKLEKKQKELGIASFGASITTEEV	835
Qy	1319	FLKVSEEDQSLENSEADVKE SRKDVLPGAEGPASGEGHAGNIARCSELTQSQASLQSASS	1378
Db	836	FLRVGK---LVDSSMDIQAIQ---LPALQ--YQHERRASDWAVDNL-----	874
Qy	1379	VGSARGDEGAGYTDVYGDYRPLFDNPQDPDNVSLQEVEAEALSRVQGSRKLDGGW-LKV	1437
Db	875	CGAMDPSDGIG-----ALIEEERTAV-----KLNTGLALHC	905
Qy	1438	RQFHGLLVKRFHCARRNSKALFSQILLPAFFVCVAMTVALSVP EIGDLPLVLSPSQYHN	1497
Db	906	QQFWAMFLKKAAYSREWKMVAQVLVPLTCVTLLALLAINYSSELFDDPMLRLTLGEY--	963
Qy	1498	YTQPRGNFIPIYANEERREYRLRLSPDASPQQLVSTFRLPSGVGATCV-LKSPANGSLGPT	1556
Db	964	-----GRTVVPFSVPGTSQLGQQ	981
Qy	1557	LNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPD EDLQAWNVS L	1616
Db	982	LS-----	983
Qy	1617	PPTAGPEMWT SAPSLPRLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILTDTG	1676
Db	984	-----EHLKDALQAEG-----QEPREVLGDL-----	1004
Qy	1677	HNVSEYLLFTSDRFLHRYG AITFG--NVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNN	1734
Db	1005	---EEFLIFRA-----SVEGGGFNERCLVAASF-----RDVGERTVVNALFNN	1044
Qy	1735	KGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQG-TDV	1793
Db	1045	QAYHSPATALAVVDNL LF-----KLLCGPHA-SIVVS NFQPRSAIQAAKDQFNEGRKGF	1098
Qy	1794	VIAIFIIVAMSFVPASFVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDMNLNYPAT	1853
Db	1099	DIALNLLFAMAFLASTFSILAVSERAVQAKHVQFVSGVHVASFWSALLWDLISFLIPSL	1158

Qy 1854 CCVILFVFDLPAYTSPTNFPVAVLSLFLLYGWSITPIMYPASFWFEVPSSAYVFLIVINL 1913  
 ::: ||: |:| : | | ||||:| |:| :|:| :|| | : |:  
 Db 1159 LLLVVFKAFDVRAFTRDGHMADTLLLLLLLYGWAIIPLMYLMNFFFLGAATAYTRLTIFNI 1218

Qy 1914 FIGITATVATFLLQLFEHDKDLKV--VNSYLKSCFLIFPNYNLGHGLMEMAYNEY----- 1966  
 || |||: :|: :: | ||: ||: || : | |  
 Db 1219 LSGI----ATFLMVTIMRIPAVKLEELSKTLDHVFLVLPNHCLGMAVSSF-YENYETRRY 1273

Qy 1967 -----INEYYAKIGQFDKMKSPFEWDI--VTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQ 2019  
 : :| | : : : | | : :| | : | : : | | : | :  
 Db 1274 CTSSEVAAHYCKKYNIQYQENFYAWSAPGVGRFVASMAASGCAYLILLFLIETNLLQRLR 1333

Qy 2020 -----RMPVSTKPVEDDQVAVSERQVRVLRGDADNDM---VKIENLTKV 2059  
 ||| : :| || || |:| | : : : | :| |  
 Db 1334 GILCALRRRRTLTELYTRMPV----LPEDQDVADERTRILAPSPDSSLHTPLIIKELSKV 1389

Qy 2060 YKSRKIGRILAVDRCLGVPRGECFLLGVNGAGKTSTFKMLTGDESTTGGEAFVNGHSV 2119  
 |: | :| |||| | | : | ||||| | ||||:| |||||:| | | :| || | :  
 Db 1390 YEQRV--PLLAVDRSLAVQKGEFCFLLGFNGAGKTTFKMLTGEESLTSGDAFVGGHRI 1447

Qy 2120 LKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVVKWALEKLELTKY 2179  
 :: :|:| :| ||| || | :| || | :| |||| : | : | | | :  
 Db 1448 SSDVGKVRQRIGYCPQFDALLDHMTGREMLVMYARLRGIPERHIGACVENTLRGLLLEPH 1507

Qy 2180 ADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIKTGRSV 2239  
 |:| | ||||| |||| | | ||||:| ||| || : : : :| : : :  
 Db 1508 ANKLVRTYSGGNKRKLSTGIALIGEPAVIFLDEPSTGMDPVARRLLWDTVARARESGKAI 1567

Qy 2240 VLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHLKNRFGDGYMITVRTKS---SQSVKDV 2296  
 :| ||||| ||||| | : :| ||| :| || : : :| : : : :  
 Db 1568 IITSHSMEECEALCTRLAIMVQGGFKCLGSPQHLKSKFGSGYSLRAKVQSEGQQEAELEEF 1627

Qy 2297 VRFFNRFPEAMKERHHTKVQYQLKSEHISLAQVFSKMEQVSGVLGIEDYSVSQTTLDN 2356  
 | : || :|:| | | | :| |:| :| :| :| |||| :| :  
 Db 1628 KAFVDLTFPGSVLEDEHQGMVHYHLPGRDLWAKVFGILEKAKEKYGVDDYSVSQISLEQ 1687

Qy 2357 VEVNFAKKQSDNLEQ 2371  
 ||:| | | :  
 Db 1688 VFLSFAHLQPPTAEE 1702

# RESULT 7

CED7\_CAEL

ID CED7\_CAEL STANDARD; PRT; 1704 AA.

AC P34358; O76287; P34359;

DT 01-FEB-1994 (Rel. 28, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE ABC transporter ced-7 (Cell death protein 7).

GN CED-7 OR C48B4.4.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM C), FUNCTION, AND MUTAGENESIS OF LYS-586;

RP GLU-639 AND LYS-1417.

RC STRAIN=Bristol N2;  
 RX MEDLINE=98297348; PubMed=9635425;  
 RA Wu Y.-C., Horvitz H.R.;  
 RT "The C. elegans cell corpse engulfment gene ced-7 encodes a protein  
 RT similar to ABC transporters.";  
 RL Cell 93:951-960(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laissster N.,  
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
 RA Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
 RA Wohldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 RN [3]  
 RP REVISIONS, AND ALTERNATIVE SPLICING.  
 RA Durbin R.;  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Functions in the engulfment of cell corpses during  
 CC embryonic programmed cell death to translocate molecules that  
 CC mediate homotypic adhesion between cell surfaces of the dying and  
 CC engulfing cells.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Comment=Experimental confirmation may be lacking for some  
 CC isoforms;  
 CC Name=c;  
 CC IsoId=P34358-1; Sequence=Displayed;  
 CC Name=a;  
 CC IsoId=P34358-2; Sequence=VSP\_000044, VSP\_000045;  
 CC Name=b;  
 CC IsoId=P34358-3; Sequence=VSP\_000044;  
 CC -!- TISSUE SPECIFICITY: Ubiquitous in embryos. Expressed in larval  
 CC germline precursors. Expression in larvae and adults is seen in  
 CC amphid sheath cells, pharyngeal-intestinal valve and phasmid  
 CC sheath cells. Low levels of expression are also seen in gonadal  
 CC sheath cells.  
 CC -!- DOMAIN: Multifunctional polypeptide with two homologous halves,  
 CC each containing a hydrophobic membrane-anchoring domain and an ATP  
 CC binding cassette (ABC) domain.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCA subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF049142; AAC24116.1; -.  
 DR EMBL; Z29117; CAA82384.2; -.  
 DR EMBL; Z29117; CAA82383.2; -.  
 DR EMBL; Z29117; CAC42271.1; -.  
 DR PIR; T42749; T42749.  
 DR WormPep; C48B4.4a; CE24856.  
 DR WormPep; C48B4.4b; CE24857.  
 DR WormPep; C48B4.4c; CE27867.  
 DR GO; GO:0016021; C:integral to membrane; NAS.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; NAS.  
 DR GO; GO:0008219; P:cell death; IMP.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 2.  
 DR ProDom; PD000006; ABC\_transporter; 2.  
 DR SMART; SM00382; AAA; 2.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 2.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
 KW ATP-binding; Transport; Transmembrane; Repeat; Glycoprotein;  
 KW Alternative splicing.

FT	TRANSMEM	24	44	POTENTIAL.
FT	TRANSMEM	256	276	POTENTIAL.
FT	TRANSMEM	306	326	POTENTIAL.
FT	TRANSMEM	335	355	POTENTIAL.
FT	TRANSMEM	436	456	POTENTIAL.
FT	TRANSMEM	963	983	POTENTIAL.
FT	TRANSMEM	1126	1146	POTENTIAL.
FT	TRANSMEM	1176	1196	POTENTIAL.
FT	TRANSMEM	1201	1221	POTENTIAL.
FT	TRANSMEM	1234	1254	POTENTIAL.
FT	TRANSMEM	1311	1331	POTENTIAL.
FT	NP_BIND	580	587	ATP (POTENTIAL).
FT	NP_BIND	1411	1418	ATP (POTENTIAL).
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	145	145	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	359	359	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	421	421	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	427	427	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	481	481	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	678	678	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	727	727	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	899	899	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	986	986	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1012	1012	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1045	1045	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1597	1597	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1632	1632	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPPLIC	496	508	Missing (in isoform a and isoform b). /FTId=VSP_000044.
FT	VARSPPLIC	992	993	Missing (in isoform a). /FTId=VSP_000045.
FT	MUTAGEN	586	586	K->R: CELL CORPSES NOT ENGULFED.
FT	MUTAGEN	639	639	E->G: CELL CORPSES NOT ENGULFED.

FT MUTAGEN 1417 1417 K->R: SOME CELL CORPSES NOT ENGULFED.  
SQ SEQUENCE 1704 AA; 191411 MW; B7502A0B24507CFE CRC64;

Query Match 12.0%; Score 1515; DB 1; Length 1704;  
Best Local Similarity 25.4%; Pred. No. 1.2e-83;  
Matches 529; Conservative 332; Mismatches 642; Indels 582; Gaps 75;

```
Qy 447 LGLLVHLM-----TSNPKILY--APAGSEVDRVILKAN-----ETFAFVGNVT 487
    || ||:|: ||:|: | | : | | : | : || :
Db 36 LGPLVYLKVKNADHTSSPENIYDNFQVKGTVEDVFLESNFIKPIYKRWCLRSDVVVGYS 95

Qy 488 HYAQVWLNISAEIRSFLEQGRLLQQLHRLWLQQYVAELRLHPEALNLSLDELPPALRQDNFS 547
    | : : | | | | | : | : | : | : | | |
Db 96 KDAAAKRTVDDLKKFAE--RFQS-----AKLKLSVKN-ESSEEQLLTVLRND--- 140

Qy 548 LPSGMALLQQLDTIDNAACGWIQFMSKVSVDIFKGFDPDEESIVNYTLNQAYQDNVTVFAS 607
    | : : | : : | : | | : | | |
Db 141 -----LPMLNETFCAINSYAAGV---VF----DEVDTVNNKKLN----- 171

Qy 608 VIFQTRKDGSLPPHVHYKIRQNSSFTEKTNEIRRAYWRPGPNTGGRF-----YFLYG 659
    | : | : | : | : | : | : | : | : |
Db 172 -----YRILLGKT-PEETWHLTETSYPYGPSSGRYSRIPSSPPYWTSA 214

Qy 660 FVWIQDMMERAIIDTFVGHDVVEPGS-----YVQMFYPCYTR-----DDFLFVIEHM 707
    | : | : | : | : | : | : | : | : |
Db 215 FLTFQHAIESSFLSS-----VQSGAPDLPITLRGLPEPRYKTSSVSAFIDFFPFI---- 264

Qy 708 MPLCMVISWVYSVAMTIQHI---VAEKEHRLKEVMKTMGLNN---AVHWVAWFITGFVQ 760
    | : : : || : : | : | : ||| : | | | | : ||
Db 265 -----WAFVTFINVIHITREIAAENHAVKPYLTAMGLSTFMFYAAHVVMAFLKFFVI 316

Qy 761 LSISVTALTALIKYGQVLMHSHVVIWLFLAVYAVATIMFCFLVSVLY---SKAKLASA 816
    | : || : : : : : : : | : : : | : : | |
Db 317 FLCSIIPLTFFVMEF----VSPAALIVTVLM--YGLGAVIFGAFVASFFNNTNSAIKAILV 370

Qy 817 CGGIIYFLSYVPYMYVAIREEVAHDKITAFEKC-IASLMSTTAFGLGSKYFALYEAVAGVG 875
    | : : || : | : | : : | : | | | : : |
Db 371 AWGAMIGISY-----KLRPEL--DQISS---CFLYGLNINGAFALAVEAISDYMRRE 419

Qy 876 IQ-WHTFSQSPVEGDDFNLLAVTMLMVDVAVYGILTWYIEAVHPGMYGLPRPWYFPLQK 934
    : : | : | : | : | : | : | :
Db 420 LNLTNMFNDSSLH---FSLGWALVMMIVDIL----- 447

Qy 935 SYWLGSG-----RTEAWEWSWPWART-----PRLSVMEEDQACAMESRRFEETR-- 979
    | : | | | : | | | : | : | : | : |
Db 448 --WMSIGALVVDHIRTSA-DFS---LRTLDFEAPEDDENQTDGVTQAQNTINEQVRNRV 501

Qy 980 ----MEEEPHTLPLV-----VCVDKLTKVYKDDKKL 1006
    || : : : | | | : :
Db 502 RRSDEIQMNPMASTSLNPPNADSDSLLEGSTEADGARDTARADIIVRNVLVKIWSTTGER 561

Qy 1007 ALNKLSLNLYENQVVSFLGHNGAGKTTTMSILTGLFPPTSGSATIYGHDIRTEMDEIRKN 1066
    | : || | | | | | | | : | : | : | | | : :
Db 562 AVDGLSLRAVRGQCSILLGHNGAGKSTTFSSIAGIIRPTNGRITICGYDVGNEPGETRRH 621

Qy 1067 LGMCPQHNVLFDRLTVEEHLWFYSRLKSMAQEEIRREMDKMIEDLELSNKRHSLVQTLSC 1126
    : || || : | : || | | | : : : : | : : | : | |
```

Db 622 IGMCPQYNPLYDQLTVSEHLKLVYGLKGAREKDFKQDMKRLLSDVKLDFKENEKAVNLSG 681  
 Qy 1127 GMKRKLSVAIAFVGGSRAILDEPTAGVDPYARRAIWDLILKYKPGRTILLSTHHMDEAD 1186  
 ||||| | : | : | : ||||| : || || : : | : : | ||||| : || : |||| :  
 Db 682 GMKRKLCVCMALIGDSEVVLLDEPTAGMDPGARQDVQKLVEREKANRTILLTTHYMDEAE 741  
 Qy 1187 LLGDRIAIISHGKLKCCGSPLFLKGTYG DG YRLTLVKRPAEPGGPQEPGLASSPPGRAPL 1246  
 ||| : | : |||| | : : || : | || || : | : | :  
 Db 742 RLGDWVFIMSHGKLVASGTNQYLKQKFGTGYLLTVV---LDHNGDK----- 784  
 Qy 1247 SSCSELQVSQFIRKHVASCLLVSDTST-----ELSYILPSEAAKKGA FERL 1292  
 || : : : | | : : || | : | | | :  
 Db 785 -----RK---MAVILTDVCTHYVKEAERGEMHGQQIEIILPE--ARKKEFVPL 827  
 Qy 1293 FQHLE-----RSLDALHLSSFGLMDTTLEEVFLKVSEEDQSLENS 1332  
 || || | : |||| ||| : || : : : : : :  
 Db 828 FQALEAIQDRNYRSNVFDNMPNTLKSQLATLEMRSFGLSLNTLEQVFITIGDK---VDKA 884  
 Qy 1333 EADVKE SRKDVLPGAEGPASGEGHAGNLARCSLTQSQASLQSASSVGSARGDEGAGYTD 1392  
 | : || : | || : | : || ||  
 Db 885 IASRQNSR-----ISHNSRNA SEPSLKPAGYDTQSSTKSA----- 919  
 Qy 1393 VYGDYRPLFDNPQDPDNVSLQEVEAEALS RVGQGSRKLDGGWLKVRQFHGLLVKRFH CAR 1452  
 | : | | : : | | : | || : : | : | : |  
 Db 920 --DSYQKLMD-----SQARGPEKSGVAKM-----VAQFISIMRKKFLYSR 957  
 Qy 1453 RNSKALFSQILLPAFFVCVAMTVALSVP EIGDLPLVLSPSQYHNYTQPRGNFIPIYANEE 1512  
 || || : || : | : : : | | | : |  
 Db 958 RNWAQLFTQVLIPIILLGL-----VGSLTTL-----KSN----- 986  
 Qy 1513 RREYRLRLSPDASPOQLVSTFRLPSGVGATCVLKSPANGSLGPTLNLSSGESRLLAARFF 1572  
 : | ||| : | : || : :  
 Db 987 -----NTDQFSVRS LTPSGIEPSKVVRWFENG TI----- 1015  
 Qy 1573 DSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPD EDLQAWNVS LPPTAGPEMWTSA PSLP 1632  
 | : | | : | :  
 Db 1016 -----PEE-----AANFE 1023  
 Qy 1633 RLVREPVRCTCSAQGTGFSCPSSVGGHPPQMRVVTGDILT DITGHNVSEYLLFTSDR FRL 1692  
 : : || : || : : | | : : |  
 Db 1024 KILRKS-----GGF-----EVLNYNTKNPL----- 1043  
 Qy 1693 HRYGAITFGNVLKSIPASFGTRAPPMVRKIAVRRAAQVFYNNKG YHSMPTYLNSLNNAIL 1752  
 | : || : | | : : : | : || : || : : |  
 Db 1044 -----PNITKSL---IGEMPPATIGMTMNSDNLEALEFN MRYYHVLPTLISMIN---- 1088  
 Qy 1753 RANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLL--QGT DVVIAIFIIVAMSFVPASF 1810  
 || | : : | : : | | | || | || : | : | : | : | : |  
 Db 1089 RARLTGTVD AEISSGVFL-----YSKSTSNSNLLPSQLIDVLLAPMLILIFAMVTSTF 1141  
 Qy 1811 VVFLVAEKSTKAKHLQFVSGCNPIIYWLANYVWDM LNYLVPATCCVILFVFDLPAYTSP 1870  
 | : || : || : : | || : | || : : | : | : | : | : | : | : | : |  
 Db 1142 VMFLIEERTCQFAHQQLTGISPITFYASLIYDGILY---SLICLI FLFMF-LAFHWMY 1197  
 Qy 1871 TNFPAVLSLFLLYGWSITPIMYPASEFWFEPSSAYVFLIVINLFIGITATVATFLL-QLF 1929  
 : | : | : | : | || | : || | || : | : | : | : | : |  
 Db 1198 DHLAIVILFWFLYFFSSVPFIYAVSFLFQSPSKANVLLI IWQVVISGAALLAVFLIFMIF 1257



Qy	1930	EHDKDLK--VVNSYLKSCFLIFPNYNLGHGEMAYNEYINEYYAKIGQFDMKSPFEWD	1987
Db	1258	NIDEWLKSILVNI FM----FLLPSYAFGSAIIT-----INTY----GMILPSEELMNWD	1303
Qy	1988	IVTRGLVAMAVEGVVGFLLTIMCQYNFLRR--PQRM PV-----STKPVEDDV----DVA	2035
Db	1304	HCGKNAWL MGTFGVCSFALFVLLQFKFVRRFLSQVWTVRRSSHNNVQPMMGDLPVCESVS	1363
Qy	2036	SERQVR LRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLG VNGAGKT	2095
Db	1364	EERERVHRVNSQNSALVIKDLTKTF-----GRFTAVNELCLAVDQKECFGLLG VNGAGKT	1418
Qy	2096	STFKMLTGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRL	2155
Db	1419	TTFNILTQGSFASSGEAMIGGRDV-TELI----SIGYCPQFDALMLDLTGRESLEILAQM	1473
Qy	2156	RGI-SWKDEARVVKWALEKLELT KYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPT	2214
Db	1474	HGFENYKAKAELI---LECVGMIAHADKLVRFYSGGQKRKISVGVAL LAPTQMIILDEPT	1530
Qy	2215	TGMDPKARRFLWNLILDLIKTGRS-VVLTSHSMEECEALCTRLAIMVNGRLRCLGSIQHL	2273
Db	1531	AGIDPKARREVWELLLWCREHSNSALMLTSHSMDECEALCSRIAVLNRGLIAIGSSQEL	1590
Qy	2274	KNRFGDGYMITVRTKSSQSVKDVVRFFNRNFP EAM LKERHHTK---VQYQL-KSEHISLA	2329
Db	1591	KSLYGNNYTMTLSLYEPNQ RDMVQVLVQTRL PNSVLKTTSTNKTLNLKWQIPKEKEDCWS	1650
Qy	2330	QVFSKMEQVSGVLGIEDYSVSQTTLDNVFNFAKKQSDNLEQQET	2374
Db	1651	AKFEMVQALAKDLGVKDFILAQSSLEETF LRLAGLDEDOLDTHST	1695

CC DOXORUBICIN RESISTANCE BY AN EXPORT MECHANISM.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M73758; AAA74717.1; -.  
 DR PIR; S27707; S27707.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR005894; Drr\_ABC\_transpt.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFAMs; TIGR01188; drrA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 KW ATP-binding; Transport; Antibiotic resistance.  
 FT NP\_BIND 41 48 ATP (BY SIMILARITY).  
 SQ SEQUENCE 330 AA; 35700 MW; 582D66C90D54E6B9 CRC64;

Query Match 3.2%; Score 405; DB 1; Length 330;  
 Best Local Similarity 30.3%; Pred. No. 3.3e-17;  
 Matches 110; Conservative 65; Mismatches 148; Indels 40; Gaps 8;

Qy 980 MEEEPHTLPLVVCVDKLTQVYKDDKKLALNKLNLNLYENQVVSFLGHNGAGKTTTMSILT 1039  
 | :|| : | ||| : |:: | ||: | || ||||:|:: :|  
 Db 1 MNTQPTR---AIETSGLVKVYNGTR--AVDGLDLNVPAGLVYGILGPNAGKSTTIRMLA 55

Qy 1040 GLFPPTSGSATIYGHDIRTEMDEIRKNLGMCPQHNVLFDRLTVEEHLWFYSRLKSMAQEE 1099  
 | | |::| :|||: :| |:: : : |: : : || |:: :  
 Db 56 TLLRPDGGTARVFGHDVTSEPDTVRRRISVTGQYASVDEGLTGTENLVMGRLLQGYSWAR 115

Qy 1100 IRREMDKMIEDLELSNKRHSLVQTLSGGMKRKLSVAIAFVGGSRRAILDEPTAGVDPYAR 1159  
 | :||: | : | |::| ||||:|: | : | : |||| |::| :|  
 Db 116 ARERAAELIDGFGLGDARDRLKTYSGGMRRRLDIAASIVVTPDLLFLDEPTTGLDPRSR 175

Qy 1160 RAIWDLI-LKYKPGRTILLSTHHMDEADLLGDRIAIISHGKLKCCGSPLFLKGTYGDG-Y 1217  
 :||:: | |::| :||| | ||||:| ||:: |: || : |  
 Db 176 NQVWDIVRALVDAGTTVLLTTQYLDEADQLADRIAVIDHGRVIAEGTTGELKSSLGSNLV 235

Qy 1218 RLTLVKRPAEPGGPQEPGLASSPPGRAPLSSCSELQVSQFIRKHVASCLLVSDTSTELSY 1277  
 || | || :: | : || :  
 Db 236 RLRL-----HDAQSRAEAERLLSAELGVTIHRD-----SDPTALSAR 272

Qy 1278 ILPSEAAKKGAFLERLFQHLERSLDALHLSSFGLMDTTLEEVFLKVS---EEDQSLNSEA 1334  
 | : | || : || | :||| : : :||| :  
 Db 273 IDDPQGMRALAELSRTHLE-----VRSFSLGQSSLDEVFLALTGHPADDRSTEEAAE 325

Qy 1335 DVK 1337  
 : |  
 Db 326 EEK 328

RESULT 9

NODI\_RHILO

ID NODI\_RHILO STANDARD; PRT; 340 AA.  
AC P23703; Q8KJI6;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Nod factor export ATP-binding protein I (Nodulation ATP-binding  
DE protein I).  
GN NODI OR MLR6164.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NZP 2213;  
RX MEDLINE=91067466; PubMed=2251131;  
RA Young C.A., Collins-Emerson J.M., Terzaghi E.A., Scott D.B.;  
RT "Nucleotide sequence of Rhizobium loti nodI.";  
RL Nucleic Acids Res. 18:6691-6691(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R7A;  
RX MEDLINE=21999272; PubMed=12003951;  
RA Sullivan J.T., Trzebiatowski J.R., Cruickshank R.W., Gouzy J.,  
RA Brown S.D., Elliot R.M., Fleetwood D.J., McCallum N.G., Rossbach U.,  
RA Stuart G.S., Weaver J.E., Webby R.J., de Bruijn F.J., Ronson C.W.;  
RT "Comparative sequence analysis of the symbiosis island of  
RT Mesorhizobium loti strain R7A.";  
RL J. Bacteriol. 184:3086-3095(2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
RT Mesorhizobium loti.";  
RL DNA Res. 7:331-338(2000).  
CC -!- FUNCTION: Part of the ABC transporter complex nodIJ (TC  
CC 3.A.1.102.1) involved in the export of LCO (lipo-chitin  
CC oligosaccharide) and a modified beta-1,4-linked N-  
CC acetylglucosamine oligosaccharide. Responsible for energy coupling  
CC to the transport system. Therefore this complex is implicated in  
CC the nodulation induction process (By similarity).  
CC -!- SUBUNIT: The complex is composed of two ATP-binding proteins  
CC (nodI) and two transmembrane proteins (nodJ) (Probable).  
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).  
CC -!- SIMILARITY: Belongs to the ABC transporter family. NodI subfamily.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration



RESULT 10

NODI\_RHIS3

ID NODI RHIS3 STANDARD; PRT; 304 AA.  
AC P72335;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Nod factor export ATP-binding protein I (Nodulation ATP-binding  
DE protein I).  
GN NODI.  
OS Rhizobium sp. (strain N33).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
OX NCBI\_TaxID=103798;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96303537; PubMed=8755627;  
RA Cloutier J., Laberge S., Prevost D., Antoun H.;  
RT "Sequence and mutational analysis of the common nodBCIJ region of  
RT Rhizobium sp. (Oxytropis arctobia) strain N33, a nitrogen-fixing  
RT microsymbiont of both arctic and temperate legumes.";  
RL Mol. Plant Microbe Interact. 9:523-531(1996).  
CC -!- FUNCTION: Part of the ABC transporter complex nodIJ (TC  
CC 3.A.1.102.1) involved in the export of LCO (lipo-chitin  
CC oligosaccharide) and a modified beta-1,4-linked N-  
CC acetylglucosamine oligosaccharide. Responsible for energy coupling  
CC to the transport system. Therefore this complex is implicated in  
CC the nodulation induction process (By similarity).  
CC -!- SUBUNIT: The complex is composed of two ATP-binding proteins  
CC (nodI) and two transmembrane proteins (nodJ) (Probable).  
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).  
CC -!- SIMILARITY: Belongs to the ABC transporter family. NodI subfamily.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U53327; AAB16898.1; -.  
DR HSSP; Q58663; 1G6H.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR InterPro; IPR005978; ABC\_transptNodI.  
DR Pfam; PF00005; ABC\_tran; 1.  
DR ProDom; PD000006; ABC\_transporter; 1.  
DR SMART; SM00382; AAA; 1.  
DR TIGRFAMs; TIGR01288; nodI; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
KW Nodulation; Transport; Membrane; Inner membrane; ATP-binding.  
FT NP\_BIND 38 45 ATP (By similarity).  
SQ SEQUENCE 304 AA; 33698 MW; 7C6A33B0364CCE14 CRC64;

Query Match 3.0%; Score 380; DB 1; Length 304;  
 Best Local Similarity 41.4%; Pred. No. 9.7e-16;  
 Matches 89; Conservative 31; Mismatches 93; Indels 2; Gaps 1;

```

Qy      2055 NLTKVYKSRKIGRILAVDRCLGLVVRPGECFGLLVNGAGKTSTFKMLTGDESTTGGEAFV 2114
          :| | || | | | : | | ||||| | ||| : : || | | : |
Db      7 DLAGVKKS--FGDKLVVNGLSFTVASGECFGLLVNGAGKSTIARMLLGMTVPDAGKITV 64

Qy      2115 NGHSLVKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKL 2174
          | | :| :| || | | | | || : : | | : | : |
Db      65 LGEPVGARSRLARKSIGVVPQFDNLDQFTVRENLLVFGRYFGMSTRKIKEVIPSLLLEFA 124

Qy      2175 ELTKYADKPAAGTYSGGNKRKLSTALIGYPAFIFLDEPTTGMDPKARRFLWNLILDLIK 2234
          | || | || || : : | || | : : ||||| : || || : | : |
Db      125 RLEKADARVGEISGGMKRRLTLARALINDPQLLVMDPEPTTGLDPHARHLIWERLRFLLA 184

Qy      2235 TGRSVVLTSHSMEECEALCTRLAIMVNGRLRCLGS 2269
          | : : : || : | || | | || || : : || |
Db      185 RGKTIILTTHFMEEAERLCDRLCVLEHGRKLAEGS 219
  
```

# RESULT 11

## NODI\_RHISN

```

ID      NODI_RHISN      STANDARD;      PRT;      343 AA.
AC      P55476;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Nod factor export ATP-binding protein I (Nodulation ATP-binding
DE      protein I).
GN      NODI OR Y4HF.
OS      Rhizobium sp. (strain NGR234).
OG      Plasmid sym pNGR234a.
OC      Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC      Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX      NCBI_TaxID=394;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97305956; PubMed=9163424;
RA      Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA      Perret X.;
RT      "Molecular basis of symbiosis between Rhizobium and legumes.";
RL      Nature 387:394-401(1997).
CC      -!- FUNCTION: Part of the ABC transporter complex nodIJ (TC
CC      3.A.1.102.1) involved in the export of LCO (lipo-chitin
CC      oligosaccharide) and a modified beta-1,4-linked N-
CC      acetylglucosamine oligosaccharide. Responsible for energy coupling
CC      to the transport system. Therefore this complex is implicated in
CC      the nodulation induction process (By similarity).
CC      -!- SUBUNIT: The complex is composed of two ATP-binding proteins
CC      (nodI) and two transmembrane proteins (nodJ) (Probable).
CC      -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC      -!- SIMILARITY: Belongs to the ABC transporter family. NodI subfamily.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
  
```

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AE000076; AAB91694.1; -.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR InterPro; IPR005978; ABC\_transptNodI.  
DR Pfam; PF00005; ABC\_tran; 1.  
DR ProDom; PD000006; ABC\_transporter; 1.  
DR SMART; SM00382; AAA; 1.  
DR TIGRFAMs; TIGR01288; nodI; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
KW Nodulation; Transport; Membrane; Inner membrane; ATP-binding; Plasmid.  
FT NP\_BIND 77 84 ATP (By similarity).  
SQ SEQUENCE 343 AA; 37917 MW; F49A7EC56E099A33 CRC64;

Query Match 3.0%; Score 380; DB 1; Length 343;  
Best Local Similarity 35.2%; Pred. No. 1.2e-15;  
Matches 93; Conservative 44; Mismatches 109; Indels 18; Gaps 4;

Qy 2006 LTIMCQYNFLRRPQMPVSTKPVEDDQVVDVASERQRVLRGDADNDM-----VKIENLTKV 2059  
: :: : | | | | | : : : : | | | | |  
Db 1 MQLLTRANVSSSPSRREPESN-----ALKQKCHGHSNADNSLSRSKSDVAIE-LTNV 50  
  
Qy 2060 YKSRKIGRILAVDRCLGVPRGECFGLLGVNAGKTSTFKMLTGDESTTGGEAFVNGHSV 2119  
| | | : | : | : | | | | | | | : : : | | | | |  
Db 51 SKS--YGDQVVVDQLSFTITSGECFGLLGPNAGKSTVSRLVLGLAPPDEGTITVLGEPV 108  
  
Qy 2120 LKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKY 2179  
: : | | | | | | | | | : : | : : : : | : |  
Db 109 PARARLARSRIGVVPQFDTLDRIFTARENLLVFGRYFGLHTRELEEAIPPLLDFALESK 168  
  
Qy 2180 ADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLIIDLIKTGRSV 2239  
| | | | | : | : | | | | | : | : | : | : : :  
Db 169 ADVPAQLSGGMQRRLTLACALINDPQLLILDEPTTGLDPHARHLIWERLRSLLALGKTI 228  
  
Qy 2240 VLTSHSMEECEALCTRLAIMVNGR 2263  
: | : | | : | | | : : : | |  
Db 229 LLTTHFMEEADRLCDRLCVIEHGR 252

#### RESULT 12

##### NODI\_RHIGA

ID NODI\_RHIGA STANDARD; PRT; 347 AA.  
AC P50332;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Nod factor export ATP-binding protein I (Nodulation ATP-binding  
DE protein I).  
GN NODI.  
OS Rhizobium galegae.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
 OX NCBI\_TaxID=399;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HAMBI 1174;  
 RX MEDLINE=99403395; PubMed=10474187;  
 RA Suominen L., Paulin L., Saano A., Saren A.M., Tas E., Lindstrom K.;  
 RT "Identification of nodulation promoter (nod-box) regions of Rhizobium  
 RT galegae.";  
 RL FEMS Microbiol. Lett. 177:217-223(1999).  
 CC -!- FUNCTION: Part of the ABC transporter complex nodIJ (TC  
 CC 3.A.1.102.1) involved in the export of LCO (lipo-chitin  
 CC oligosaccharide) and a modified beta-1,4-linked N-  
 CC acetylglucosamine oligosaccharide. Responsible for energy coupling  
 CC to the transport system. Therefore this complex is implicated in  
 CC the nodulation induction process (By similarity).  
 CC -!- SUBUNIT: The complex is composed of two ATP-binding proteins  
 CC (nodI) and two transmembrane proteins (nodJ) (Probable).  
 CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. NodI subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X87578; CAA60881.1; -.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR005978; ABC\_transptNodI.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFAMs; TIGR01288; nodI; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 KW Nodulation; Transport; Membrane; Inner membrane; ATP-binding.  
 FT NP\_BIND 81 88 ATP (By similarity).  
 SQ SEQUENCE 347 AA; 38435 MW; AC791210C44C9A6C CRC64;

Query Match 3.0%; Score 379; DB 1; Length 347;  
 Best Local Similarity 31.3%; Pred. No. 1.4e-15;  
 Matches 103; Conservative 56; Mismatches 130; Indels 40; Gaps 6;

Qy 2019 QRMPVSTKPVEDDVDVASERQVRVLR-----GDADNDMVKIENLTKVYKSRKIGR 2067  
 :| : | : | : || | | : : : | :| : :  
 Db 6 EREMLRPKTIAMDQNSASARSNPEREIKTGRLEPASNSAPTMAIDLQAVTMIYRDKTV-- 63  
 Qy 2068 ILAVDRLCLGVRPGECFGLLGVNGAGKTSTFKMLTGDESTTGGEAFVNGHSLVKELLQVQ 2127  
 || | ||| ||||| ||||| : : || | : : | : | | : :  
 Db 64 ---VDSLSEFGVRAGECFGLLGPNAGAGKSTITRMLLGMATPSAGKISVLGLPVPKGKARLAR 120  
 Qy 2128 QSLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTY 2187  
 |:| | | | | ||:| : : | : : : || :| ||



Db 121 ASIGVVSQFDNLDMEFTVRENLLVFGRYFQMSTRAIEKLIPSLLEFAQLEAKADVRSVL 180

Qy 2188 SGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLIIDLIKTGRSVVLTSHSME 2247  
 ||| ||:|: | ||: | : |||||:| || :| : |: |:::|:| |:

Db 181 SGGMKRRLTLARALVNDPQLLILDEPTTGLDPPARHQIWERLRSLLIRGKTILLTTHMMD 240

Qy 2248 ECEALCTRLAIMVNGRLRCLG-SIQHLKNRFG-----DGymITVR-----TKS 2289  
 | | :| || :: ||: | : :: | | : || |

Db 241 EAERMCDRLCVLEGGRMIAEGPPLSLIEDIIGCPVIEVYGGNPDELSLIVRPHVDRIETS 300

Qy 2290 SQSV-----KDVVRFFNRFPEAMLKER 2312  
 :: | || | || | ||

Db 301 GETLFCYTVNSDQVRAKLREFPSLRLLER 329

# RESULT 13

## NDI2\_RHIME

ID NDI2\_RHIME STANDARD; PRT; 335 AA.

AC Q8GNH6;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Nod factor export ATP-binding protein I (Nodulation ATP-binding protein I).

GN NODI.

OS Rhizobium meliloti (Sinorhizobium meliloti).

OG Plasmid megaplasmid.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.

OX NCBI\_TaxID=382;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=102L4;

RX MEDLINE=22343004; PubMed=12455608;

RA Barran L.R., Bromfield E.S., Brown D.C.;

RT "Identification and cloning of the bacterial nodulation specificity gene in the Sinorhizobium meliloti-Medicago laciniata symbiosis.";

RL Can. J. Microbiol. 48:765-771(2002).

CC -!- FUNCTION: Part of the ABC transporter complex nodIJ (TC 3.A.1.102.1) involved in the export of LCO (lipo-chitin oligosaccharide) and a modified beta-1,4-linked N-acetylglucosamine oligosaccharide. Responsible for energy coupling to the transport system. Therefore this complex is implicated in the nodulation induction process (By similarity).

CC -!- SUBUNIT: The complex is composed of two ATP-binding proteins (nodI) and two transmembrane proteins (nodJ) (Probable).

CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).

CC -!- SIMILARITY: Belongs to the ABC transporter family. NodI subfamily.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

```

DR      EMBL; AF522456; AAN62904.1; -.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR003439; ABC_transporter.
DR      InterPro; IPR005978; ABC_transptNodI.
DR      Pfam; PF00005; ABC_tran; 1.
DR      ProDom; PD000006; ABC_transporter; 1.
DR      SMART; SM00382; AAA; 1.
DR      TIGRFAMS; TIGR01288; nodI; 1.
DR      PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR      PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW      Nodulation; Transport; Membrane; Inner membrane; ATP-binding; Plasmid.
FT      NP_BIND        69      76      ATP (By similarity).
SQ      SEQUENCE      335 AA;  36878 MW;  8826A6330FD63CC6 CRC64;

```

Qy	1982	SPFEWDIVTRGLVAMAVEGVGFLLTIMCQYNFLRRPQRMVSTKPVEDDVVASERQRV	2041
		:  :  :     :  :  :  :	
Db	12	SPFEW----KGDAGPSVKTL-----RPHAIPSA-----IDLAS-----	41
Qy	2042	LRGDADNDMVKIENLTKVYKSRKIGRILAVDRLCLGVRPGECFGLLGVNGAGKTSTFKML	2101
		:     : :                    : : :  :	
Db	42	-----VTKSYGDKPV-----VDGLSFTVAAGECFGLLGPNGAGKSTITRMI	82
Qy	2102	TGDESTTGGEAFVNGHSVLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWK	2161
		:       : :             :  :  :  :	
Db	83	LGMTTPATGVITVLGVPVPSRARLARMGIGVVPQFDNLDSEFTVRENLLVFGRYFRMSTR	142
Qy	2162	DEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKA	2221
		:   :               :  :    :       :	
Db	143	EIEAVIPSLLEFARLENKVDARVSDLSGGMKRRLTLARALINDPQLLILDEPTTGLDPHA	202
Qy	2222	RRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNGR	2263
		:  :   :  :  :             :	
Db	203	RHLIWERLRSLLARGKTILLTTHIMEEAERLCDRLCVLEAGR	244

## RESULT 14

```

ID      _NDI1_RHIME          STANDARD;          PRT;      355 AA.
AC      052618;
DT      30-MAY-2000 (Rel. 39, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Nod factor export ATP-binding protein I (Nodulation ATP-binding
DE      protein I).
GN      NODI OR RA0472 OR SMA0864.
OS      Rhizobium meliloti (Sinorhizobium meliloti).
OG      Plasmid pSymA (megaplasmid 1).
OC      Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC      Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX      NCBI_TaxID=382;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=1021;

```

RX MEDLINE=21396509; PubMed=11481432;  
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,  
 RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,  
 RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,  
 RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,  
 RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;  
 RT "Nucleotide sequence and predicted functions of the entire  
 RT Sinorhizobium meliloti pSymA megaplasmid.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).  
 RN [2]  
 RP SEQUENCE OF 143-355 FROM N.A.  
 RC STRAIN=1021;  
 RA Barnett M.J., Long S.R.;  
 RT "Nucleotide sequence of nodIJ region of Rhizobium meliloti pSymA.";  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Part of the ABC transporter complex nodIJ (TC  
 CC 3.A.1.102.1) involved in the export of LCO (lipo-chitin  
 CC oligosaccharide) and a modified beta-1,4-linked N-  
 CC acetylglucosamine oligosaccharide. Responsible for energy coupling  
 CC to the transport system. Therefore this complex is implicated in  
 CC the nodulation induction process (By similarity).  
 CC -!- SUBUNIT: The complex is composed of two ATP-binding proteins  
 CC (nodI) and two transmembrane proteins (nodJ) (Probable).  
 CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. NodI subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE007237; AAK65130.1; -.  
 DR EMBL; AF043118; AAB97762.1; -.  
 DR PIR; H95320; H95320.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR005978; ABC\_transptNodI.  
 DR Pfam; PF000005; ABC\_tran; 1.  
 DR ProDom; PD0000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFAMs; TIGR01288; nodI; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 KW Nodulation; Transport; Membrane; Inner membrane; ATP-binding; Plasmid;  
 KW Complete proteome.  
 FT NP\_BIND 89 96 ATP (By similarity).  
 SQ SEQUENCE 355 AA; 39268 MW; 4DC8696D98C335DC CRC64;

Query Match 2.9%; Score 365.5; DB 1; Length 355;  
 Best Local Similarity 33.8%; Pred. No. 9.5e-15;  
 Matches 95; Conservative 35; Mismatches 102; Indels 49; Gaps 4;

QY 1982 SPFEWDIVTRGLVAMAVEGVVGFLLTIMCQYNFLRRPQRMVSTKPVEDDDVDVASERQRV 2041  
 ||||| | | |: | || :|||

```

Db      32 SPFEWKDQTGLAVKTAIPG-----AKPTV-AIDVAS----- 61
Qy      2042 LRGDADNDMVKIENLTKVYKSRKIGRILAVDRCLCLGVRPGECFGLLGVNGAGKTSTFKML 2101
          :|| | : : : : | | | | | | | | | | :| :
Db      62 -----VTKSYGDKPV-----INGLSFTVAAGECFGLLGPNAGKSTITRMI 102
Qy      2102 TGDESTTGGEAFVNGHSLKELLQVQQSLGYCPQCDALFDELTAREHLQLYTRLRGISWK 2161
          | : || | | | : : | | | | | | | | :| :
Db      103 LGMTTPGTGEITVLGVFPVPSRARLARMRIGVVPQFDNLDLEFTVRENLLVFGRYFRMSTR 162
Qy      2162 DEARVVKWALEKLELTKYADKPAGTYSGGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKA 2221
          : | : || | | | | | | | | | : | | | | :| :
Db      163 EIEAVIPSLLEFARLENKADARVSDLSGGMKRRLTLARALINDPQLLILDEPTTGLDPHA 222
Qy      2222 RRFLWNLILDLIKTGRSVVLTSHSMEECEALCTRLAIMVNG 2262
          | :| : | : | : : : | | | | | | : : |
Db      223 RHLIWERLRSLLARGKTILLTTHIMEEAERLCDRLCVLEAG 263

```

# RESULT 15

## YBHF\_ECOLI

```

ID   YBHF_ECOLI          STANDARD;          PRT;   578 AA.
AC   P75776; Q9R7S3; Q9R7S4;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update).
DE   Hypothetical ABC transporter ATP-binding protein ybhF.
GN   YBHF OR B0794 OR SF0744 OR S0785.
OS   Escherichia coli, and
OS   Shigella flexneri.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562, 623;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [3]

```

RP SEQUENCE FROM N.A.  
 RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;  
 RX MEDLINE=22272406; PubMed=12384590;  
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
 RA Yu J.;  
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
 RT through comparison with genomes of Escherichia coli K12 and O157.";  
 RL Nucleic Acids Res. 30:4432-4441(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;  
 RX MEDLINE=22590274; PubMed=12704152;  
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
 RA Schwartz D.C., Blattner F.R.;  
 RT "Complete genome sequence and comparative genomics of Shigella  
 RT flexneri serotype 2a strain 2457T.";  
 RL Infect. Immun. 71:2775-2786(2003).  
 CC -!- SIMILARITY: Belongs to the ABC transporter family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE000181; AAC73881.1; ALT\_INIT.  
 DR EMBL; D90716; BAA35454.1; ALT\_INIT.  
 DR EMBL; D90717; BAA35460.1; ALT\_INIT.  
 DR EMBL; AE015103; AAN42379.1; ALT\_INIT.  
 DR EMBL; AE016980; AAP16256.1; ALT\_INIT.  
 DR EcoGene; EG13314; ybhF.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 2.  
 DR ProDom; PD000006; ABC\_transporter; 2.  
 DR SMART; SM00382; AAA; 2.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
 KW Hypothetical protein; ATP-binding; Transport; Repeat;  
 KW Complete proteome.  
 FT DOMAIN 1 237 ABC\_TRANSPORTER\_1.  
 FT DOMAIN 330 559 ABC\_TRANSPORTER\_2.  
 FT NP\_BIND 40 47 ATP (POTENTIAL).  
 FT NP\_BIND 362 369 ATP (POTENTIAL).  
 FT CONFLICT 44 44 A -> E (IN REF. 4).  
 SQ SEQUENCE 578 AA; 63132 MW; DB3B3FA213490F3C CRC64;  
  
 Query Match 2.8%; Score 354; DB 1; Length 578;  
 Best Local Similarity 15.1%; Pred. No. 1.1e-13;  
 Matches 194; Conservative 112; Mismatches 236; Indels 744; Gaps 18;

Qy	990	VVCVDKLTkVYKDDKkLALNKLSlnLYENQVVSFLGHNGAGKTTTMSILTGLFPPTSGSA	1049
		: ::     :     :   ::   :             :	
Db	5	VITLNGLEKRFPgMDKPAVAPLDCTIHAGYVTGLVGPdGAGKTTLMRMLAGLLKPDsgSA	64
Qy	1050	TIYGHDIRTEMDEIRKnlGMCPQHnVLFDRlTVEEHLWFYSRLKsMAQEEIRREMDKMIE	1109
		:     :         ::           :     :   :   :   : : ::	
Db	65	TVIGFDPIKNDGALHAVLGYMPQKfGLYEDLTVMENlnLYADLRsvTGEARKQTFARLLE	124
Qy	1110	DLELSNKRHSLVQTLsGGMKRKLSVAIAfVGGsRAIILDEPTAGVDPYARRAIWDLILKY	1169
		:       : :                 :     : : :	
Db	125	FTSLGPFTGRLAGKLSGGMKQKLGLACTlVGEpKVllLlDEPGVGVDpISrRELWQMvHEL	184
Qy	1170	K-PGRTILLSTHhMDEADLLGDRIAIISHGKLKCCGSPLfLKgTYGDGYRLTLVKRPAEP	1228
		: :     : : :   :	
Db	185	AGEGMLILWSTSYLDEAEQCRD-VllMNEGELLYQGEp-----	221
Qy	1229	GGPQEPGLASSPPGRAPLSSCSELQVSQfIRKHVASCLLVSDTSTELSYILPSEAaKKGa	1288
Db	222	-----	221
Qy	1289	FERLFQHLERSLDALHLSSfGLMDTTLEEVFLKVSEEDQsLENSEADVKEsRKDVLPGAE	1348
Db	222	-----	221
Qy	1349	GPASGEGHAGNLARCSELtQSQASlQSASSVGSARGDEGAGYTDVYGDYRPLFDNPQDPD	1408
		:     :   :	
Db	222	-----KALTQTMa-----GRSF-----LMTSPH---	239
Qy	1409	NVSLQEVEAEALSrVQGQSRKLDGGWLKVrQFHGLLVKRFHCARRNSKALfSQILLPAFF	1468
		:   :	
Db	240	-----EGNRKL-----	245
Qy	1469	VCVAMTVALSVPEIGDLpPLVLSPsQYHNYTQPRGNfIPYANEERREYRLRLSPDASPQQ	1528
		:     :   :     : :       :   :	
Db	246	----LQRALKLPQVSD-----GMIQgKSVRLILKKEATPDD	277
Qy	1529	LVSTFRLPSGVGATCVLKSPANGSLGPTlNLSSGESRLLAARFFDSMCLESFTQGLPLSN	1588
		:	
Db	278	I-----RHADGM-----	284
Qy	1589	FVPPPPSPAPSDSPASPDedLQAWNvSLPPTAGPEMWTsAPSLPRLVREPVRCTCSAQGT	1648
		:	
Db	285	-----PEI-----	287
Qy	1649	GFSCPSSVGCHPPQMRVVTGDILTdITGHNVSEYLLFTSDRfRLHRYGAITFGNVLKSIP	1708
		: :     :	
Db	288	-----NINE-----TTPRFE-----	297
Qy	1709	ASFGTRAPPMVRKIAVRRAAQVFYNNKGYHsMPTYLNSLNNAILRANLPKsKGNPAAYGI	1768
		: : :	
Db	298	-----DAFIDLLGGA-----	307
Qy	1769	TVTnhPMNKTSASLSLDYLLQGTdVVIAIfIIvAMSfVPASfVVfLVAEKSTKAKHLQfV	1828
Db	308	-----	307

Qy	1829	SGCNPIIYWLANYVWDMLNYLVPATCCVILFVFDLPAYTSPTNFFAVLSLFLLYGWSIT	1888
		:	
Db	308	-----GTSES	312
Qy	1889	PIMYPASFWFEVPSSAYVFLIVINLFIGITATVATFLLQLFEHDKDLKVNSYLKSCFLI	1948
		:	
Db	313	PL-----	314
Qy	1949	FPNYNLGHGLMEMAYNEYINEYYAKIGQFDKMKSPFEWDIVTRGLVAMAVEGVVGFLITI	2008
		:	
Db	315	-----GAILHTVEGTPG-----	326
Qy	2009	MCQYNFLRRPQRMFVSTKPVEDDDVVASERQQRVLRGDADNDMVKIENLTKVYKSRKIGRI	2068
		: : :	
Db	327	-----ETVIEAKELTK-----KFGDF	342
Qy	2069	LAVDRLCLGVRPGECFGLLVNGAGKTSTFKMLTGDESTTGGEAFVNGHSLVKELLQVQQ	2128
		:   :                 : :     : :     : :	
Db	343	AATDHVNFAVKRGEIFGLLGPNGAGKSTTFKMMCGLLVPTSGQALVLGMDLKESSGKARQ	402
Qy	2129	SLGYCPQCDALFDELTAREHLQLYTRLRGISWKDEARVVKWALEKLELTKYADKPAGTYS	2188
		:   :     : :   : : : :   : :	
Db	403	HLGYMAQKFSLYGNLTVEQNLRFSGVYGLRGRAQNEKISRMSEAFGLKSIASHATDELP	462
Qy	2189	GGNKRKLSTAIALIGYPAFIFLDEPTTGMDPKARRFLWNLIIDLIKTGRSVVLTSHSMEE	2248
		:   :   :   :         :       : :   :   :	
Db	463	LGFKQRLALACSLMHEPDILFLDEPTSGVDPLTRREFWLHINSMVEKGVTVMVTTHFMDE	522
Qy	2249	CEALCTRLAIMVNGRLRCLGSIQHLK	2274
		: :   :	
Db	523	AE-YCDRIGLVYRGKLIASGTPDDLK	547

Search completed: September 1, 2004, 10:53:18  
 Job time : 54 secs